



Evaluation of the COGITO system

Andersen, V.; Andersen, H.H.K.

Publication date:
2002

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Andersen, V., & Andersen, H. H. K. (2002). *Evaluation of the COGITO system*. Risø National Laboratory. Denmark. Forskningscenter Risø. Risø-R No. 1363(EN)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Evaluation of the COGITO system

V. Andersen and H.H.K. Andersen

**Risø National Laboratory, Roskilde
August 2002**

Abstract *This report covers the deliverable D7.2 of the COGITO project. It presents the evaluation of an 'intelligent' agent integrated into an e-commerce system. The aim of the agent is to support the user partly through direct communication and partly by guiding him/her for navigating the web-site and for benefiting efficiently from the functionality of the site. In order to make the report a stand-alone report, it contains an introduction including 'Study of end-user needs and behaviour' exposing problems concerning e-commerce in general (taken from the D6.1), and a brief review of the objectives for the COGITO project in order to clarify the aims of the project. However, the important part of the report presents the methodology for evaluating the performance of the COGITO agent and the outcome of this evaluation. The questionnaire for eliciting the background information for the test persons and their personal responses to the performance of the agent is presented in appendices.*

Contents

1	Introduction.....	4
1.1	Problems concerning e-commerce in general.	4
1.1.1	Surfing and buying experience.....	4
1.1.2	Transfer of traditional shopping behaviour	4
1.1.3	Degree of involvement in the product	5
1.1.4	Lack of trust.....	5
1.1.5	Non-transparent interfaces.....	5
1.1.6	Buying products	6
1.1.7	Delivery	6
1.1.8	Summary	7
1.2	Objectives of the COGITO project	7
2	Methods.....	9
2.1	User requirements specification method	9
2.2	Evaluation method.....	10
2.3	Evaluation test sessions.....	12
2.4	Background of respondents.....	12
3	Conversation log analysis	17
4	Analysis of eye-tracking data.....	22
5	Questionnaire results.....	26
5.1	The outcome of the evaluation for each scale	27
6	Conclusion	32
7	References.....	33
	Appendix 1: List of tasks to be solved by the groups of test persons using the BOL avatar	36
	Appendix 2: List of tasks to be solved by the groups of test persons using the COGITO agent	38
	Appendix 3: Background information of the test persons.....	40
	Appendix 4: Questionnaire for the impression of the test persons related to the COGITO agent/BOL site	42
	Appendix 5: Introduction to users	53

1 Introduction

1.1 Problems concerning e-commerce in general.

The process of buying products and services on the Internet often implies a high degree of complexity and uncertainty about the conditions of information seeking, about items for sale, the purchase of wanted products and the actual navigation on a site. Some important problems concerning e-commerce in general and shopping at Internet bookstores in particular are outlined below.

1.1.1 Surfing and buying experience

A problem that is not only related to e-commerce but to the use of the Internet in general is getting people started on the Web surfing and more specifically making their first purchase. Many people still haven't acquired experience using the Internet and feel insecure about the technology and the potential benefits. Many elderly people's lack of experience with computers is a strong barrier to the exploitation of e-commerce sites and there is a need for a willingness to make decisions and get involved in a new area of interest.

1.1.2 Transfer of traditional shopping behaviour

Using traditional metaphors for shopping on web sites, for example reassembling the physical environment, can create advantages as well as problems. The obvious advantages are the possibility of recalling known patterns thereby making decision-making easier. On the other hand users transfer their expectations concerning shopping to the electronic media and can experience uncertainty when confronted with the differences that, no matter the resemblance with traditional shops, lie within the Internet way of shopping. Users are so to speak forced to make their model of shopping fit into a web structure with which they are not familiar.

In physical shops, customers are guided along specific shelves or arrangements in the shop and thereby exposed to certain categories of products. This is not only a strategy for marketing but it also functions as the customer's shopping lists – a sort of reminder of needed items. Problems arise because the e-stores seldom provide this feature and the users feel they are not led around the shop in an effective manner. They experience that they forget to buy things or miss important offers thereby being forced to use extra time to return to the site to finish shopping.

1.1.3 Degree of involvement in the product

There is more interest in online-buying products with a low degree of involvement, that is, a low enthusiasm about the product. One example is groceries. Users are motivated for buying this kind of products on the Internet because it is a relief of their shopping burden and because the products are identical to the known products at the local store. Problems arise when e-commerce sites are concerned with products with a high degree of involvement, for example clothes and to some degree also books. Customers feel a need for touching the fabric, finding the right size, feeling the quality and weight of books, comparing different items etc. The physical sense of touching the product and the visual representation are of utmost importance to the decision of buying or not buying. Many users find it unnatural to relate to products in the form of text and they don't feel inclined to buy these products on the web.

1.1.4 Lack of trust

Internet stores frequently offer several opportunities to the users to get tailored recommendations for books and other items, create personal lists, and obtain news from the shops. In order to do so users must submit information about their personal preferences, email addresses, and in the case of buying items, a credit card number is imperative. Two problems arise from the need for submission of information. One concerns the creation of a user profile and another the handling and utilisation of the given information. Some people are not interested in being confronted with possibilities for action that is based on information about earlier choices or interactions. They feel that this could constrain their action possibilities even though they might in fact enjoy the personalisation of the shopping experience provided on the basis of a user profile. Concerning the handling of information, users often state a need for information about which of their data are saved and for what purpose. They fear that information could be sold to other organisations and that they then would receive a massive amount of junk mail from both the specific store and other unknown companies. Users are generally cautious about submitting their email addresses and feel an even greater uncertainty about submitting their credit card number. This uncertainty is enhanced if it is not made clear who the owner of the site is or if detailed information about the product is not available. Trust in that a specific e-commerce site/organisation handles the user information properly and respects the privacy of the individual buyers is therefore paramount to the utilisation of the given offers. Consequently, trust grows from experience supporting this policy of the e-commerce organisation.

1.1.5 Non-transparent interfaces

An obvious problem that's often encountered at the Internet is the lack of well-arranged sites. The main site of the tested e-bookstore doesn't display a clear structure of the site and its contents. It is messy and overloaded with information. This results in difficulties navigating the site - in particular locating wanted functions or information. Because of this design, users are forced to scroll even though this way of exploring a

site is tiring. Furthermore, it is difficult to obtain a comprehensive view of the site and confusion about action possibilities arises.

The following illustrates the difficulties users experience trying to locate a specific function on the tested website. When wanting to do combinatorial searches, for instance combining a specific author with a certain subject, the users know they want an advanced search but they aren't able to locate the proper function. Instead, they get stuck in the simple search function placed on the top of the page just below the navigation bar. The advanced search is actually placed directly above the other search possibility in the navigation bar under the heading "search". This means that it is very easy to mistake the simple search window for being a part of this heading and not explore this option further. In this way the problem of ill-structured web sites is also a question of not separating and making functions clear to the users and not providing enough explicative information.

Another problem is the structuring of information on the site. Often information is categorised in a hierarchical structure in which subjects that are more detailed are contained in superior classes. This can make it difficult to find subjects if the higher classes are not known or the contents is not explained thoroughly enough.

1.1.6 Buying products

Users sometimes seem to find the buying procedure difficult. Problems can arise due to the attempt of the bookstore to highlight the buying function, which in this case is placed close to other information about the actual item. Instead of leading the buyer to the buying function, users interpret a highlighted area as a commercial banner that has nothing to do with the book purchase, and they don't pay any attention at all to the contents of the area. This can result in attempts to find the buying function in other parts of the site forcing the respondents to navigate around the site often getting caught up in dead ends from which it is difficult to find one's way back to the starting point. These kind of fundamental problems inhibit the current sale of products and surely discourage the buyers from returning to the site.

1.1.7 Delivery

The issue of delivery is also a substantial problem. For example, uncertainty about how, where and when the bought items will be delivered. Furthermore, the dependency on the opening hours of the local post office or having to be at home at specific hours represent a great obstacle to using Internet stores. A topic in close relation to the delivery issue is that in many cases extra costs are included because of the collection of products, for instance from a supermarket, and the delivery itself. This can make the price of buying on the Internet higher than buying from the physical stores and make the motivation for electronic shopping decrease.

1.1.8 Summary

The common theme for the mentioned problems is uncertainty. Uncertainty about the new media, the new ways of shopping, adequate representation of products, trust in the e-commerce sites, the navigation of specific sites and the actual procedures for buying. It is very important to overcome these problems in order to facilitate the use and acceptance of e-commerce.

1.2 Objectives of the COGITO project

The rapid evolution of interactive Internet services has led to both a constantly increasing number of modern Web sites, and to an increase in their functionality, which, in turn, makes them more complicated to use. Thus, any attempt to enhance the consumer-supplier relationship in e-commerce has to meet the challenge of coping with two almost contradictory goals: A useful e-commerce application should not only mimic traditional catalogues, order forms and other printed material which used to be the basis of communication between consumers and suppliers. Instead, the inherent potential for interactive data processing and man-machine dialogue should be used by e-commerce applications to meet the user's need for immediate situation-specific response, instantly available problem-specific advice, and better ways to access and inspect the supplier's offer. However, the currently prevailing graphical user interfaces, which rely on menu selection and navigation, require a considerable cognitive overhead. This may be tolerable to frequent users, but will in many cases deter casual users, especially those who are not yet used to computers. Hence, we need to combine the usefulness of a value-added service with a high degree of usability, and dedicated measures to build up trust and confidence in inexperienced users.

To meet these conditions the interaction must be, at the same time, as natural as possible, thus enabling users to rely on their communicative skills, it must convey precise and relevant information, and address the personal background of the individual user. The interface must use best practice solutions to achieve a high degree of dialogue intelligence, and employ an appropriate graphical design.

The solution we propose for project COGITO is based on "intelligent personalized agents" which represent virtual assistants or advisors (also visually) by modelling their ability to support customers. There are many possible applications for such virtual assistants. They could instruct customers in the use of a Web site, point out new offers, help sift through products, and other support. There have already been some efforts made in developing chat robots ("chatterbots") based on expert systems.

A chatterbot is a software system capable of engaging in conversation (in written form) with a user, often entertaining the user with some "smalltalk" – sometimes accompanied by cartoons expressing emotions. In most applications, chatterbots are used as guides who can show the user around on a Web site. This can be a stereotyped "guided tour" allowing only few deviations; however, this concept has to be abandoned when the Web site is too large to be explored by navigation, or contains

too many offers. This is the case in e-commerce applications, where pages are generated on demand by retrieving data from a product database and assembling the result into HTML pages, usually hit lists of searches.

Virtual assistants must be capable of flexible behavior if they are to be acceptable to users on a long-term basis. Simple chatterbots, such as the first system of this type, ELIZA, and most of its successors (see <http://bots.internet.com/search/s-chat.htm> for an overview) only simulate conversation without utilizing any knowledge about the individual users and their actual behavior during online sessions. Such simple chatterbots are not powerful enough to serve as a medium for customer advice. This means that, in addition to some of the abilities already available (e.g., help question answering controlled by simple event-action rules), a further reaching dialogue management will be needed to help accomplish two major goals. First, in order to achieve an adequate, non-stereotypical repertoire of reactions, the individual dialogue situation must be interpreted; and second, dialogues that are more complex allow goal-directed strategies to be pursued (cooperative behavior, convincing argumentation). This kind of dialogue intelligence will be based on elaborated dialogue rules allowing the system to interpret a wide variety of situations that may occur. Dedicated editor tools will support the construction of these rule sets.

Whereas an increase in general dialogue intelligence can be achieved by elaborate rule sets, the naturalness of the dialogue depends on the degree in which the system is able to adapt to individual users, whether it is able to learn about their preferences and attitudes during the dialogue, and memorize them for later use. For this purpose, we will include learning mechanisms that extract permanent features of a given user from the dialogue (of course, the user must consent to this, and will be given an opportunity to inspect and change the data). The resulting user profiles will be further analyzed to automatically extract usage patterns from the data given about user communities. This helps content providers to tailor their offers to the customers' needs, and can be used to generate assumptions about new users, when they start to converse with the system. Published research to date shows that a further development of personalized interfaces into more flexible dialogue-oriented interfaces could increase the acceptance of such personalized agents.

While important, the increased ergonomic usability and personalization of chatterbots are only first steps. The main problem of most of today's Web services is that they offer manifold navigation options and (usually simple) search functions, but leave it up to users to find their way through the many interface functions, understand them and interrelate them cognitively. Usually, users have to decide themselves which sequence of actions must be performed to solve a given task. Complex search queries, for example, must be constructed step by step. Beginners and occasional users are often daunted by the complexity of today's services and thus need "pro-active" support or advice from the system in order to fully utilize the range of functions available. Therefore, we will add a component for intelligent access to the supplier's repository, which will act as a "prompter" helping the chatterbot in problematic retrieval

situations (too many, too few hits, etc.). It will rely on a repository of search heuristics, and exploit the profiles as well as domain knowledge provided by the content manager. The latter will be capable to harvest a supplier's XML-based Web site and extract structural and semantic information. As any automatic assistance must be limited, we will add a gateway to the supplier's call center. This again should contribute to increase the consumer's trust and confidence.

The expressive visualization of a virtual advisor – e.g., as an animated cartoon “Persona” – can be a direct and useful complement to the proposed dialogue approach. By being able to take the initiative, rather than simply reacting to user input and commandos, a system can take on the role of an independent agent during dialogue. To make this role as a true counterpart transparent, it is helpful to visualize the agent – thus the agent is also visually present and can go beyond the communicated content to express functional aspects of its dialogue contributions (questions, recommendations, warnings, etc.) by means of mimic and gestures. Moreover, the Persona can also visually express emotional aspects of interaction (interest, warmth, humor), thus contributing significantly to a relaxed atmosphere and increased attractiveness of the service.

In order to verify the assumptions underlying the design decisions above, and to find out appropriate ways to adjust the system parameters, the technical development will be accompanied and heavily influenced by in-depth evaluations of both the individual components as well as the system as whole. The last part of the project will be devoted to exploitation efforts as well.

2 Methods

2.1 User requirements specification method

In order to facilitate the evaluation of the COGITO outcome, the user requirements were presented categorised in terms of means-end relations (see D7.1, ‘Study of the end-user needs and behaviour’; Andersen et al, 2001). The framework for this categorisation is shown in the Figure 1, presenting by the middle column the hierarchy in general and by the left-hand column the means-end hierarchy in a condensed form utilising fewer levels in the hierarchy. This presentation has proved sufficient and successful for specification of user requirements, and was utilised in the COGITO project, indicating the strategic goals on the highest level, procedures supporting these goals at the next lower level, and – at the lowest level – the operations from which these procedures are created. In this representation, each level will be specified by the next upper level concerning the reason or background for an action, and by the next lower level concerning how this action may be supported (see the right-hand column).

User Requirements	Means- End	Relations among levels
Strategic requirements	Goals and constraints	Why
	Abstract Functions	Why What
Procedural requirements	Generalized Function	Why What How
	Physical functions	What How
Operational requirements	Physical Form	How

Figure 1: Means-end relations presented in general form in the middle column, in squeezed form utilised for specifying user requirements for COGITO at the left hand column, and with indication of the relations among the various levels at the right hand column.

2.2 Evaluation method

System evaluation normally constitutes three levels of evaluation procedures:

- verification, which is a check of implementation of operations specified in the user requirements, and therefore directly related to the lower level of the hierarchical representation of user requirement, the operational requirements;

- evaluation, which is a check of the presence of the functionality specified in the user requirements, i.e. is the system capable of executing all the sequences of operations needed for fulfilling the goals specified in the requirements. This part of the test procedure is directly related to the middle part of the hierarchical presentation of the requirements, the procedural requirements;

- last, but not least, the validation takes care of – based on user satisfaction – to test whether the system is of any value to the end users, i.e. do they perform better, more efficient and with a higher success rate than without having the system available. The question here is the difference between developing the system right, i.e. following carefully all the elicited requirement specifications, or developing the right system, i.e. a system that really is of benefit to the end users.

The logical way of evaluating a system is to take the top-down approach in which the evaluation and validation is tested by user interaction with the system. In case this test does not end up satisfactorily, the next step would be to take the bottom-up approach starting with the verification phase checking the implementation of operational features and continuing with check of the functional features.

So, the COGITO system evaluation is based on the top-down approach resting on system to user interaction. In more detail the ‘evaluation and validation’ of the COGITO agent has been performed by letting groups of test persons solve various tasks related to searching general information or specific products utilising the agent and its linking with

specific BOL sites related to the questions and wishes of the users. The evaluation is partly based on quantitative measures, such as, e.g., the length of sentences of the users indicating a real conversation in contrast to using just a search-engine technique, the stereotypical use of the sentences utilised by the agent, and the number of fall back sentences indicating a missing interpretation of the request from the user. Likewise, objective and quantitative measure of eye-tracking specifying the time the user spent looking at the agent, the answers given by the agent, or the BOL site itself. Furthermore, the evaluation is qualitatively based on the users' subjective assessment of using the system and of the outcome of their search. This is done through direct interviews and through fulfilment of detailed questionnaires enlightening the general impression and understanding of the agent when the agent reacted to requests from the user and presented the suggestions through effective links to the BOL site.

So, during the session the visual perceptions of the test persons are monitored, and following the session the eye-tracking data and the log files of the communication between the user and the agent have been analysed to check, respectively, the visual attraction of various parts of the screen and the communicative performance of the agent related to various supporting features. The eye movements were monitored using a headset free eye tracking equipment in order not to distract the test person and thereby influence the performance. Finally, the test persons are requested to complete the questionnaire revealing their satisfaction with the system and the agent concerning various aspects, such as impression, control, effectiveness, navigability, learnability, aidability, and comprehension of the agent.

In order to have a reference for evaluating the COGITO proactive agent, a baseline session was performed using the BOL site equipped with a 'BOL avatar' for comparing the two agents. This agent had a level of chatting performance in line with existing agents of today and was integrated to the BOL site by having simple links to products being requested by the customer, i.e. this agent had no proactive features.

In order to test - for various levels of users - the benefit from the agent experienced by the users related to their previous experience from using the net, the test group was divided into two groups, one holding novices in using the net and one holding experienced users.

After being introduced to the system the test person was asked to deal with a number of tasks prepared for the test session. The tasks included problems like getting an overview about BOL, may be a guided tour of the site, find a specific book or books about a specific topic, or it could be search for information about how to order and pay, or possibly about the security in using credit cards. The complete list of tasks for this round is shown in appendix 1 and 2, respectively, for the sessions using the BOL avatar and the COGITO proactive agent. The type of tasks for the two sessions is the same, except for the books suggested for a search is more IT related for the last session in order to benefit from the domain in which the pro-activity is mostly developed in this version of the COGITO agent.

The questionnaires, one for revealing the background information about the test person him/herself and one for having the personal indication of the test person regarding various features of the agent and the site, partly concerning the overall feelings of the system related to the expectations, and partly more detailed concerning expectations, impression, effectiveness, aidability, learnability, and control and navigation of the system, is found in appendix 3.

2.3 Evaluation test sessions

Four groups of eight persons each were recruited for the test sessions, two groups of novices and two groups of experienced users in order to test the validity of the COGITO prototype for each of these types of end-users. Even though we do know that groups of eight persons are not representative for the population as a whole, this is the size of group within the reach related to the time and money available for the COGITO evaluation, and furthermore, still reasonable for getting a first impression of the functionality and validity of the COGITO agent.

The evaluation sessions were performed as follows:

1. Baseline session performed 22-23 November 2001 at Osnabrück University
 - a. Novices using the BOL web-site equipped with the BOL avatar
 - b. Experienced users using the BOL web-site equipped with the BOL avatar
2. Final evaluation session performed 31 January and 1 February 2002 at Osnabrück University
 - a. Novices using the final prototype including the proactive agent, version January 2002
 - b. Experienced users using the final prototype including the proactive agent, version January 2002

2.4 Background of respondents

The test groups were recruited via the Osnabrück University, and therefore included quite a few students, even though it was attempted to have a reasonable distribution between students and employed persons. This was mainly obtained for the two test groups related to the basis sessions and to some degree concerning the group of novices for the real COGITO agent session, whereas the group of experienced persons for this session involved students only (see Figure 2). However, due to the fact that the average age of the groups were all in the middle of twenties (see Figure 3), and nearly all the participants were educated persons, this is not expected to be of vital importance. For the sex distribution, the groups are rather equal in males and females (see Figure 4), except for the Nov.2 group having a surplus of females. In general, this group acted a bit more unsatisfied than the other groups. If this is related to the surplus of females is outside the scope of this report.

The criteria for being placed in a group of novices or experienced users are related to the experience concerning use of computers as well as experience in using the facilities on the Internet or using the net for buying purposes. It is clearly seen from the illustration of 'Computer experience of users (Figure 5), that the experienced users have nearly the double or more of computer experience as compared with the less experienced ones, here called novices even though they all had reasonable experience. Likewise, it is seen from the graphs that the experienced users have a more frequent use of the Internet at home. In addition, they even use the Internet during working hours (see Figure 6). In addition, they have a higher variety in the facilities utilised on the Internet (see Figure 7). Finally, the experienced users have much more experience in buying various items via the Internet than the 'novices', see Figure 8, who was very limited in varying their purchase if buying at all (see Figure 9).

The previous experience with agents on the Internet was not seen as an important criterion, and the background of the test persons showed that no one the novices had any experience with agents, whereas a few of the experienced users had worked with agents one or more times (see Figure 10).

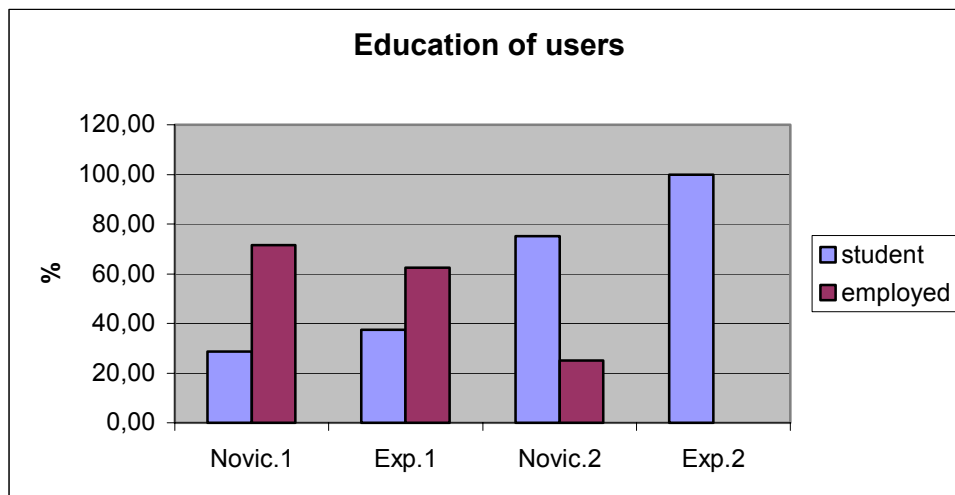


Figure 2 Users divided according to employment

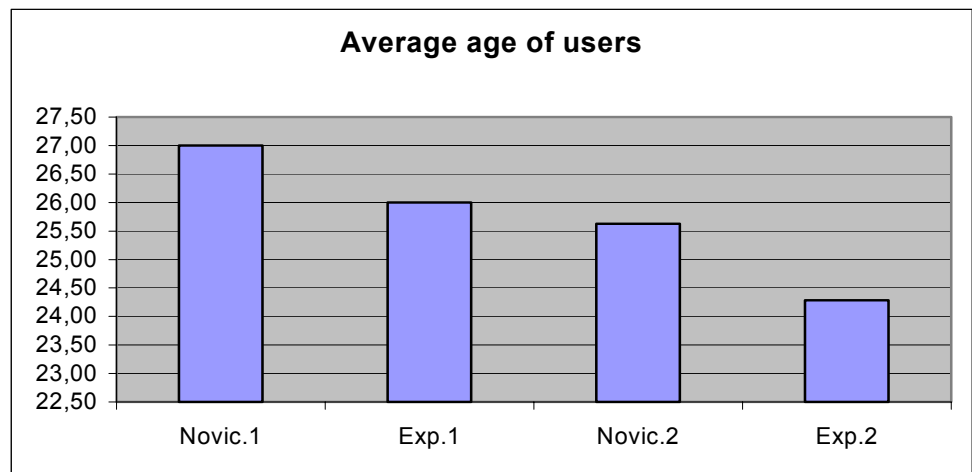


Figure 3 Average ages of users

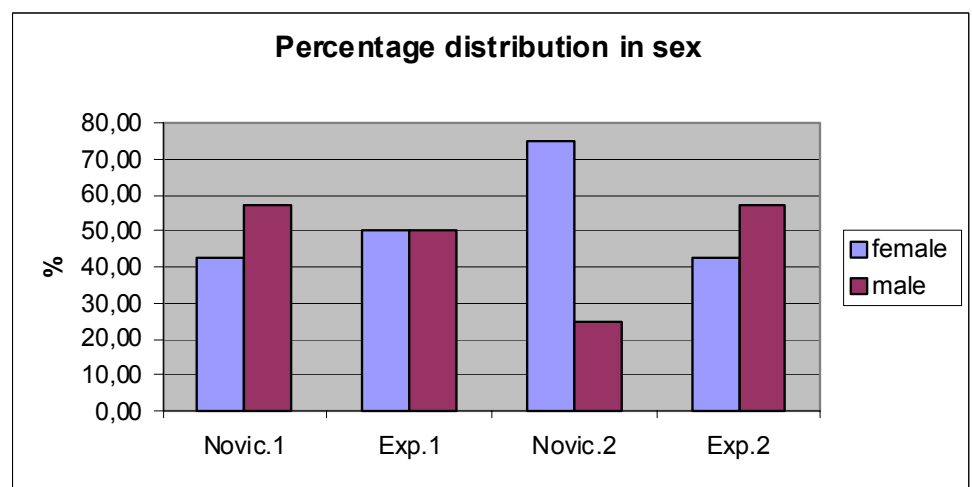


Figure 4: Distribution of users according to sex

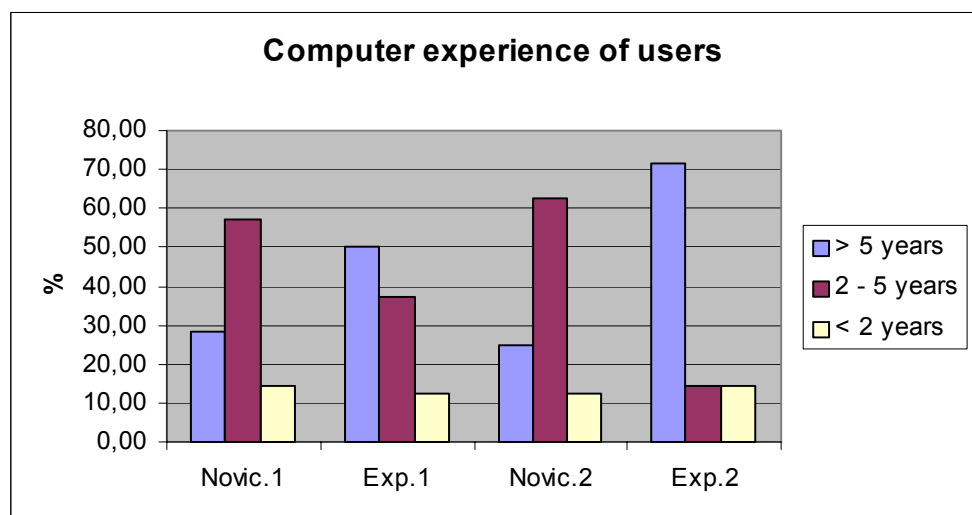


Figure 5: Computer experience of users

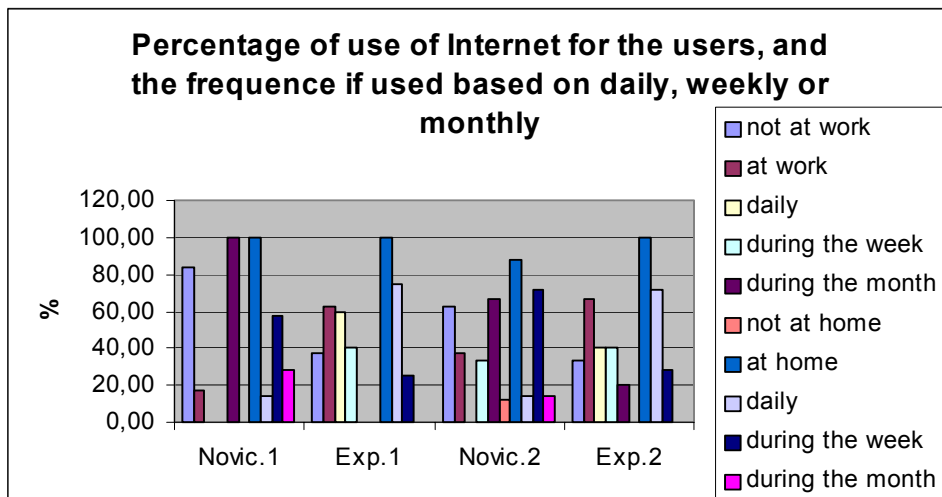


Figure 6: Use of Internet

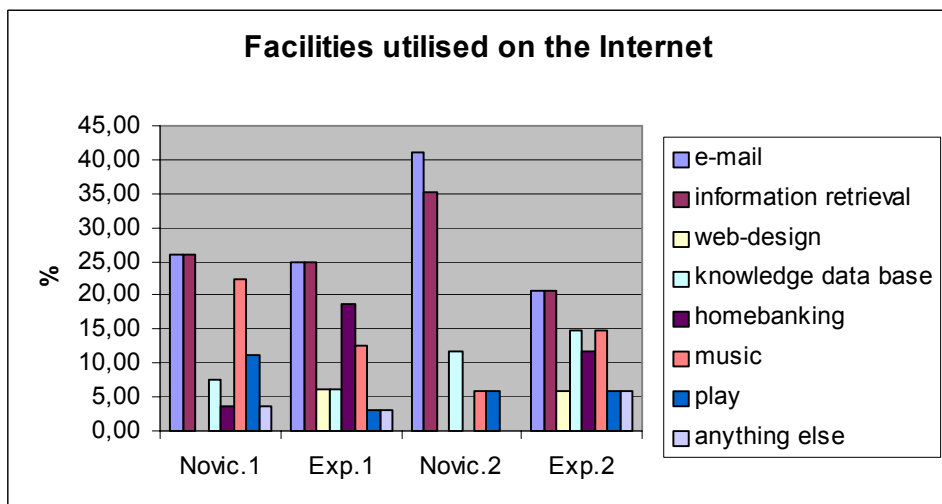


Figure 7: Use of facilities

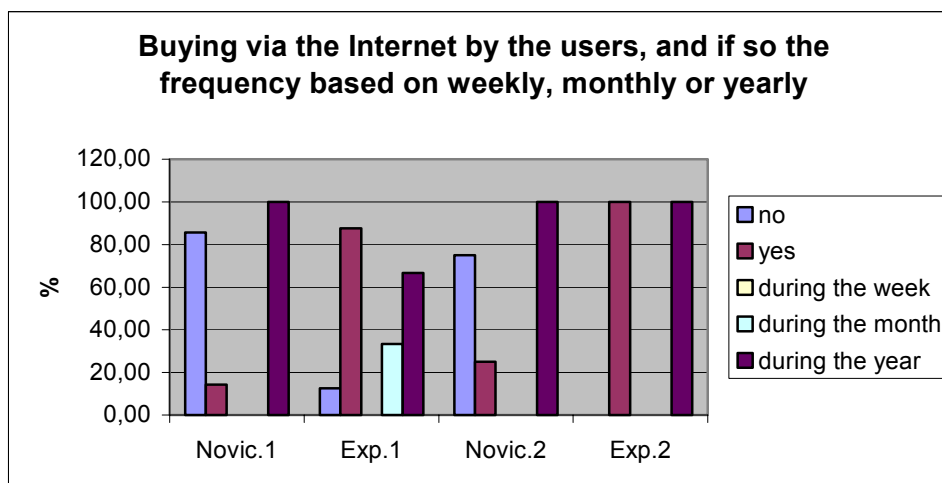


Figure 8: Buying via the Internet

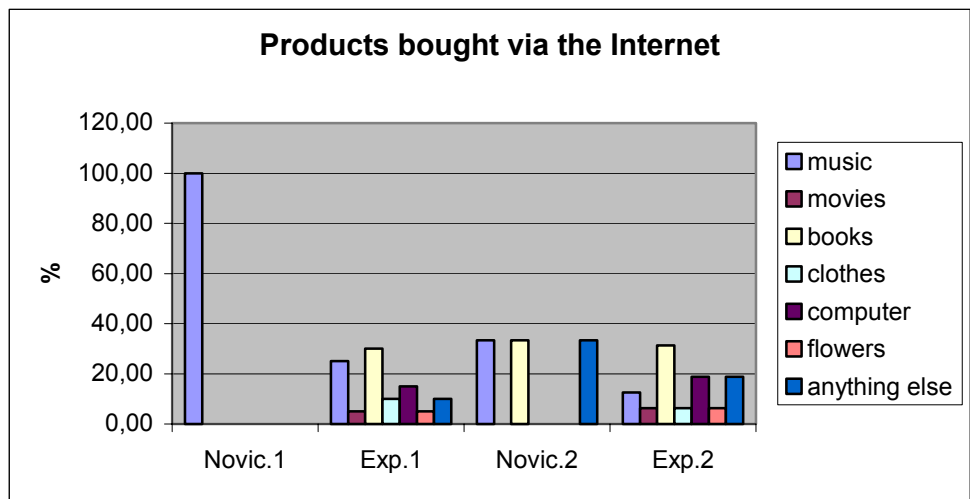


Figure 9: Items bought via the Internet

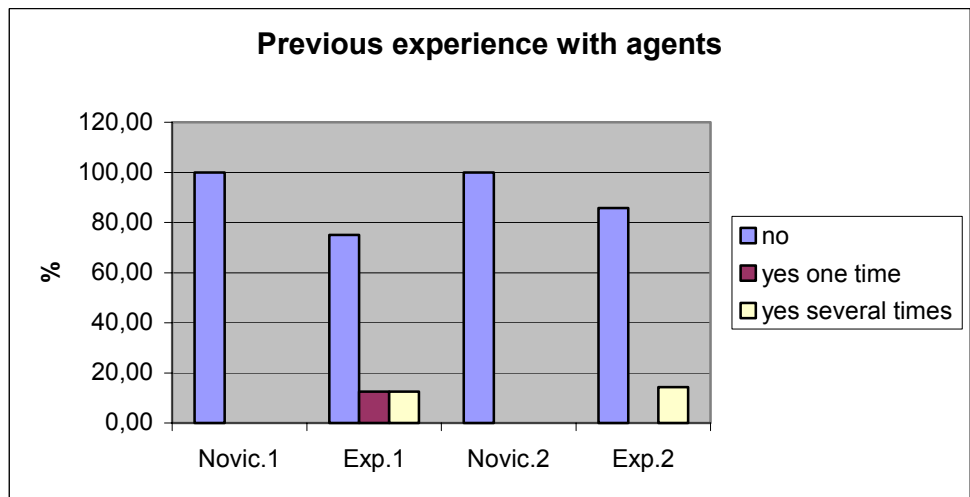


Figure 10: Experience with agents

3 Conversation log analysis

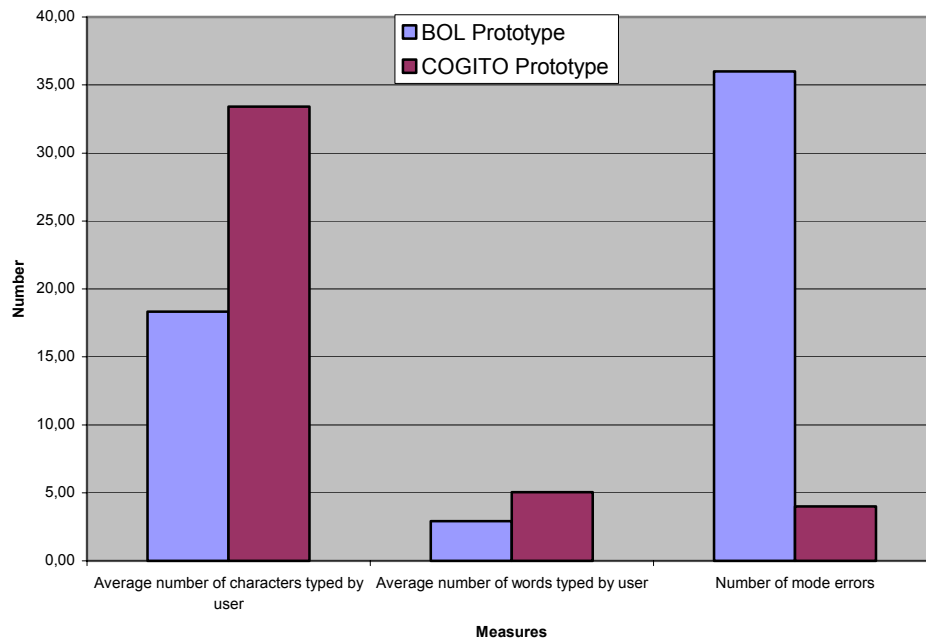


Figure 11: The number of characters and word typed by subjects for both agents. In addition, the number of “mode errors” is shown.

As seen in the conversation log analysis (Figure 11) the users of the COGITO prototype typed approximately twice the amount of characters typed by users of the BOL prototype. On an average a COGITO user query contained 5,05 terms while a BOL user query contained 2,95 terms. In an analysis of queries posed by users at Excite, a major Internet search service, Jansen et al (2000) found that web queries are short. On an average, a query contained 2,21 terms. The BOL agent users used queries that were more than 30% longer, while the COGITO queries on an average contained more than twice the number of terms (app. 120% more) compared with the Excite users and more than 70% more than the BOL users. Nevertheless compared to studies of queries in traditional information retrieval systems like online databases (e.g. DIALOG) and public access catalogues (e.g. library catalogues) both the BOL and the COGITO queries are significant shorter. In traditional information retrieval systems the queries on an average varies from 7 to 15 terms dependent on the users expertise varying from novices to very experienced (for more information on this topic see e.g. Fenichel (1981) and Spink and Saracevic (1997)).

In fact, hand one the main idea of introducing intelligent agents on the web is exactly to overcome some of the obstacles of traditional information retrieval like e.g. the use of Boolean operators and in the same way to allow users in natural way to type their queries in a

conversational manner. With an average query length of 5,05 terms, the pro-active COGITO agent tended to perform better than the traditional web based search engines (like the ones on Excite) without demanding the users to use any Boolean operators, and as shown in Figure 12 with relatively good performance in terms of search results. In addition, the pro-activeness of the COGITO agent seems to urge people to type in more words than the more “passive” BOL agent where users tended to use the agent as a traditional search machine.

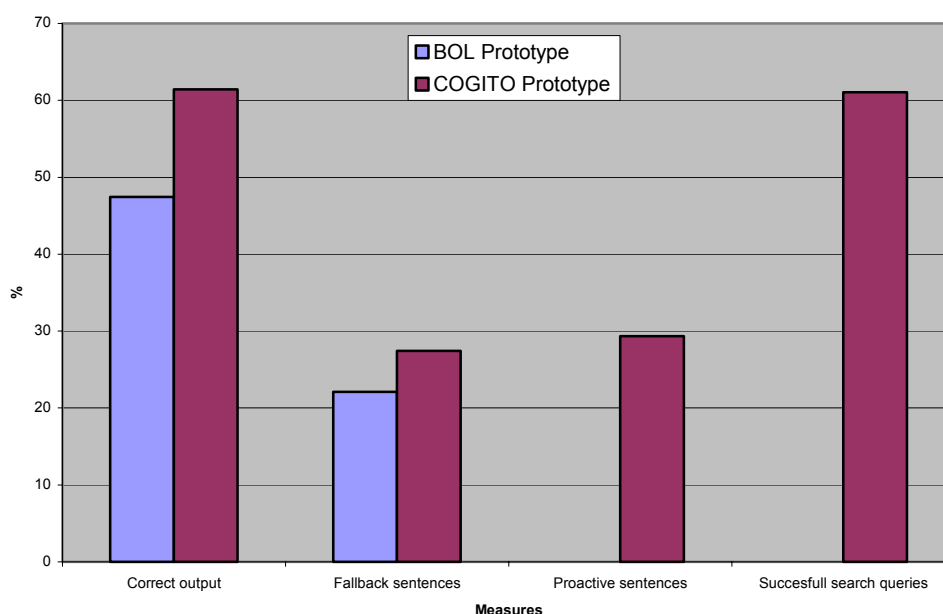


Figure 12: The result of the analysis of agent-user conversation logs in terms of number of correct text output, fallback sentences, and pro-active sentences. In addition, the figure shows the agent performance with respect to number successful search queries

Table 1: Analysis of log files for the two prototypes. All numbers are averages shown in percentages.

	Correct output	Fallback sentences	Pro-active sentences	Successful search queries
BOL Prototype	47%	22%	0%	0%
COGITO Prototype	61%	27%	29%	61%

The analysis of the conversation log served the purpose to measure the conversation performance in terms of number of correct text output, fallback sentences, and pro-active sentences (see Figure 12). The measure “Correct text output” is based on manual analysis and interpretation of successful elements of the agent-user dialogue consisting of one user text

input string, e.g. a request for information or a search query, and one agent output text string, e.g., delivering a correct answer and/or requesting further information from the user. Here below is an example from the conversation log that we interpret as a one correct output from the agent.

User: Please tell me about you selection of books? (Was für Bücher sind im Angebot?)

Agent: bol.de offers an almost complete selection of German books? (Der Buch-Shop von bol.de bietet nahezu alle lieferbaren deutschsprachigen Bücher?) (Quote from BOL agent conversation logs; Eds. translation.)

The COGITO agent shows a better performance than the BOL agent with respect to the correct output category (61% vs. 47%). It seems that the COGITO agent is better at recognizing search terms; it has a better vocabulary and rule set than the BOL agent. Furthermore, the fact that the BOL agent is relatively passive and therefore does not in the same sense as the COGITO agent produce true pro-active requests in response to user input adds further to the explanation. A pro-active agent sentence is counted every time the agent successfully takes a word or a string of words and constructs a contextually meaningful response to user input. Here below is an example of such a pro-active conversational sequence:

User: Please tell me something about Microsoft Office? (Was können Sie mir zum Office Paket sagen?)

Agent: Which Office software do you most often use? (Welches Programm aus dem Office Paket nutzen Sie am häufigsten?) (Quote from COGITO agent conversation logs; Eds. translation.)

The BOL agent of course also asks questions, but they are more passive and general, and the agent does not utilise user input in stating further requests. These types of sentences we do not count as proactive. One example is given below:

User: What can I do online (Was kann ich online machen)

Agent: How can I be of further assistance to you (Womit kann ich Ihnen sonst weiterhelfen) (Quote from BOL agent conversation logs; Eds. translation.)

We have also analysed the heterogeneousness (or homogeneity) of the conversations using two measures: the proportions of fallback sentences and the various sentence categories applied by the agents. We consider a large occurrence of fallback sentences as an expression of poor conversation performance. That is, there is a risk that the users will get frustrated if the agent “speaks” in set phrases or clichés, which means that the user experiences a rather stereotypical conversation. Here below is a couple of examples of fallback sentences:

Agent: Interesting expression. Never heard it before. What do you really mean by that? (Interessante Formulierung. Habe ich so noch nie gehört.

Was Sie meinen denn damit?) (Quote from COGITO agent conversation logs; Eds. translation.)

and,

Agent: Please reformulate your input? (Bitte formulieren Sie Ihre Eingabe noch einmal neu?) (Quote from BOL agent conversation logs; Eds. translation.)

In this sense, both agents did not show optimal performance. The BOL agent performed a bit better than the COGITO agent (22% vs. 27%), but one has no feeling of a significant difference between the two. In addition, the COGITO agent showed a larger variation in the conversation by using 9 different categories of fallback sentences while the BOL agent only used 6 different categories. We also measured the proportion of different sentence categories. The BOL agent applied 66% more sentences than the COGITO agent, but only produced 28% more different categories of sentences. That is, relatively the COGITO agent performed approximately 30% better than the BOL agent using this measure did.

We have also looked at the proportion successful search queries generated by the COGITO agent. In this respect a successful query is counted every time the agent on the basis of selected user input prompts the bol.de search machine with queries that produce a correct list of search results in terms of relevance for a given user situation (task) no matter if the user recognizes this. That is, we have repeated all the search queries listed in the COGITO agent conversation logs using the bol.de search machine and analysed the result in relation to the users tasks. Our stop rule was 2 pages of results. This is more than most users would go through. As shown by Jansen et al. (2000) most users in their study (58%) did not access any result past the first page. Both the conversation logs and the videos show that this is also the case for the users in this study. This fact might influence our satisfaction measures based on the questionnaires. The non pro-active BOL agent does not produce search queries. Instead, it operates with a concept of static “deep linking” based on general input from the user. However, most static deep links did not function because BOL changed its platform during the evaluation session.

We also measured what we have named “mode errors”. During the evaluation of the BOL agent, it became clear that if the user, for example, asked for a book, then afterwards - or if the search failed - the agent did not leave this book search mode. An example might help to clarify what we mean:

User: How are the terms for ordering books. (ich moechte gern ein buch bestellen wie sehen die bedingungen aus)

Agent: I will be pleased to help you in your search. May I inspire you! Do you know the title of the book or a part of it? If yes, please tell me and I will try to find the book for you. If not, please type “No” so I can help you further. (Ich helfe Ihnen gerne bei der Suche. Lassen Sie sich von mir inspirieren! Wissen Sie den Titel des Buches oder Teile davon? Wenn ja, sagen Sie ihn mir bitte und ich suche das Buch für Sie.

Ansonsten geben Sie bitte 'Nein' ein, damit ich Ihnen anders bei der Suche helfen kann).

User: Is it safe to pay by credit card? (ich moechte mit kreditkarte bezahlen ist dies auch sicher?)

Agent: On the basis of this title I have found the following books. Is your book among these? "The agent links to a list of books that includes credit card in the title. (Eds)" (Folgende Bücher konnte ich zu Ihrem Titel finden. Ist das dabei, was Sie gesucht haben?) (Quote from BOL agent conversation logs; Eds. translation.)

The user ask for the tems of ordering a book and the agent takes the word "book" as a signal that the user want s to search for a book. The user though wants to get information about the term for ordering not to search for a specific book. Next, the agent prompts the user to answer "Yes" or "No" but apparently the user have misunderstood the out of context answer from the agent. He does not follow the agent's procedure for handling book searches and is now more specific in asking for the safety of using a credit card. The agent interprets this input as a title of a book and provides a list of books that includes the term "credit card" in the title.

As seen in Figure 11 the COGITO agent performed much better than the BOL agent with respect to this measure. Only 4 mode errors was found compared to the BOL agent's 36. In addition the COGITO agent's mode errors seem differ in their nature from those of the BOL agent. Below is an example from the COGITO agent conversation logs:

User: What books about programming languages can I get? (Welche Buecher ueber Programmier-Sprachen sind erhaeltlich?)

Agent: Can I be of any further assistance to you, Tina? (Darf ich noch etwas für Sie tun, Tina?)

User: I am interested in books about java or html! Ich interessiere mich fuer Buecher ueber java oder html!

Agent: Can I be of any further assistance to you, Tina? (Was kann ich noch für Sie tun, Tina?)

User: Do you have books about programming languages? (Gibt es Buecher ueber Programmier-Sprachen?)

Agent: Can I be of any further assistance to you, Tina? (Was kann ich noch für Sie tun, Tina?) (Quote from COGITO agent conversation logs; Eds. translation.)

The agent keeps asking the same question repeatedly. It seems like the agent is not able to extract any search terms from the user input. Since we only found 4 mode errors in the analysis of the conversation logs and that these did not seem to be systematic, we suspect that they may be due to minor technical prototype problems. The BOL agent mode errors on the other hand are very systematic. It is not possible to "jump" from one sub-level category in the rule base to another category. Notice the exclamation mark in user input 2 in the quotation from the COGITO agent conversations logs. There is no doubt that mode errors can create extensive confusion and in the end frustration among users.

4 Analysis of eye-tracking data



Figure 13 shows a snapshot of the BOL site and agent as they appeared during the evaluation of the BOL agent 19-20th November 2001. The red colour “boxes” (dark for black and white copies) mark the examined areas of interest.

We have used a SMI remote eye-tracking system to measure the respondent’s visual behaviour during the evaluation session (see Figure 16 to view the set-up.). This device is non-intrusive. The respondents can behave as they normally would in front of a computer display. The eye-tracking system samples the eye-movements at 50 Hz to a data file. In addition we have video recorded the eye-movements together with the graphic signal from the computer. The data is sampled only during respondents task solving. That is, we have recorded data from the point in time where the respondent is finished reading the task situation out loud and until she/he has finished the task, gives up or is stopped by the moderator. We have divided the screen into 5 so-called “Areas Of Interest” (AOI’s, see Figure 13 and Figure 14) and named them

1. The agent torso, which shows the animation of the agent
2. The agent text output field, where text from the agent is displayed
3. The user input field, where the user can type, e.g., requests for information
4. The BOL site where, e.g., search results are displayed
5. The right lower corner, that shows the background of the screen.



Figure 14 shows a snapshot of the BOL site and agent as they appeared during the evaluation of the COGITO agent 31st January to 1st February 2002. The red colour “boxes” (dark for black and white copies) mark the examined areas of interest.

We have then calculated the amount of visual attention paid to each of these AOI's in percentage of all viewing time during the task situations (see Figure 15).

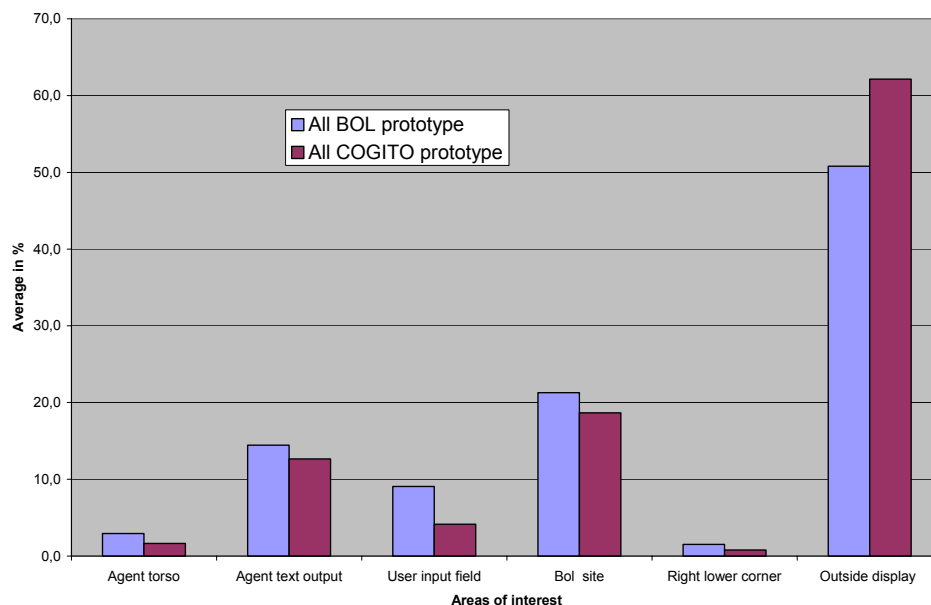


Figure 15 shows the result of eye-tracking analysis for all respondents for the two agents. Numbers are averages in percentage.

The smallest amount of viewing time has been spent looking at nothing at the right corner of the display. This is rather common to normal viewing behaviour when people is either daydreaming or solves cognitive task that do not require any new visual input. In addition, a small amount of viewing time has been spent on the visualisation of the agents. More visual attention has been paid to the BOL agent visualisation compared to animation the COGITO agent (2,9% vs. 1,6%). The BOL agent animation has attracted approximately double as much visual attention than the COGITO agent. This is probably due to the more the photo-like appearance, the obliging attitude, and a larger repertoire of gesticulations compared to the cartoon-like appearance of the COGITO agent.

With respect to the agent part of the system, most viewing time has been spent at the text output field. This is not surprising that it takes time to read the text. In addition, some user had to scroll back to read longer paragraphs since the text “ran” to fast. There is only a 1,8 % (14,4% vs. 12,6%) difference between the two prototypes with respect the text output AOI. A larger difference is measured in terms of the user input field (9,1% vs. 4,1%). It seems like the BOL prototype respondents used more time in checking their keyboard strokes than the COGITO respondents. This result is even more distinctive because of the average number of words typed by the BOL respondents compared to the COGITO respondents (see also Figure 11). The BOL agent’s larger amount of mode errors and output errors may also play a role. That is, the BOL respondents wanted to be sure not to make any “typos.” making the moderator think that the agent errors were their fault. Another explanation is, that the two groups differ with respect to their typing abilities. The analysis of the profiles of the two groups seems not to indicate any difference on this matter (see also Section 2.4).

With respect to visual attention at the bol.de site there is only a little difference between the two prototypes (21,3% vs. 18,6%). On explanation could be that most BOL agent deep links did not function well enough, which more often lead the BOL respondents to use the BOL site on its own to deal with a given situation than was the case with the COGITO respondents. In addition, the fact that the BOL agent deep links some times lead to rather surprising pages could play a role. If a respondent e.g. asked on how to use a credit card she/he was linked to a page that promoted a wide range of colourful cheap Christmas offers.

In general, all groups of respondents used approximate half of their display viewing time on the agent and half on the bol.de site. The BOL group of respondents looked at the site 21,3% of their display viewing time and 26,4 % at the agent. The COGITO group looked at the site 18,6 of their display viewing time and 18,4 % at the agent.

For both prototypes, much viewing time has been spent outside the display. This may not be surprising since the agent requires input in terms of written text using the keyboard. Not looking at the display means that we have no eye-data. This means that either the respondent blinks, eye-data is lost while the respondent look at the screen due to a less optimal calibration or the person looks outside the display at the keyboard, at the task description or at the moderator. From the video of

the respondents, it is clear that approximately 40% of the viewing time outside the display is spent on typing at the keyboard.

In both prototypes, the agent and the bol.de site are split into separate windows. This is a somewhat artificial separation due to the fact that BOL could not give the developers access to their real operative web place. Nevertheless, if the agents have been incorporated into the site the eye-tracking study will most probably have shown the same results.

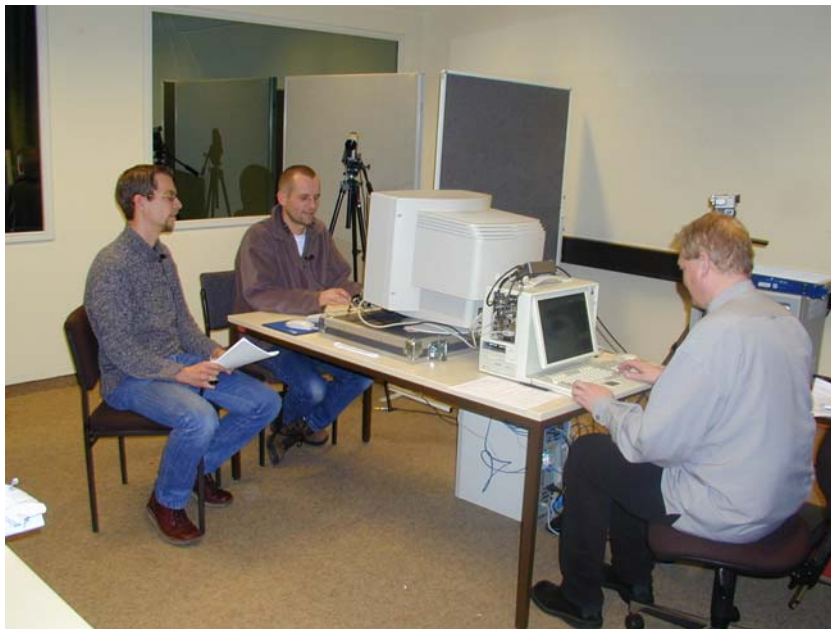


Figure 16: Photo of set-up: Test person, moderator, and at the right hand side monitoring of the eye-tracking calibration.



Figure 17: Overview of monitoring equipment like the working environment including eye-tracking indication, and video surveys and two observers.

5 Questionnaire results

The result of the user satisfaction questionnaire is partly based on superior questions related to the BOL site and the agent, relatively, and in more detail dealing with questions related more directly to the agent. See the questionnaire for both parts of questions in appendix 3.

For the first part of the questionnaire, the questions have been formulated in a quick de-briefing form to be completed by the answers, 'very unsatisfied', 'unsatisfied', 'satisfied', or 'very satisfied'. For the more detailed part the questions have deliberately been formulated changing in positive or negative terms in order to force the respondents to be very cautious in understanding and responding to the questions. This will prevent the respondents from a mechanical indication of satisfaction related to right or left side of the field for responding to the questions. The potential responses to these questions were 'strongly disagree', 'disagree', 'agree', or 'strongly agree'. These responses were later interpreted to the same terms as mentioned above, i.e. 'very unsatisfied', 'unsatisfied', 'satisfied', or 'very satisfied'.

The result of the 'quick de-briefing' indicates a rather high satisfaction concerning the BOL site as such, in fact with very high agreement among the four groups, especially when summing the result related to 'very unsatisfied' and 'unsatisfied', and similar for 'satisfied' and 'very satisfied'. For all the groups the indication is around 17% unsatisfied and 83% satisfied users. Concerning the 'quick de-briefing' for the agent the result is somewhat more differentiated among the groups with a higher satisfaction by the experienced users than by the novices for both agents. Furthermore, the shift from the state of the art agent to the COGITO agent indicates a slight increase in satisfaction for both novices and experienced users. See the 'Quick de-briefing', Figure 18 and Figure 19.

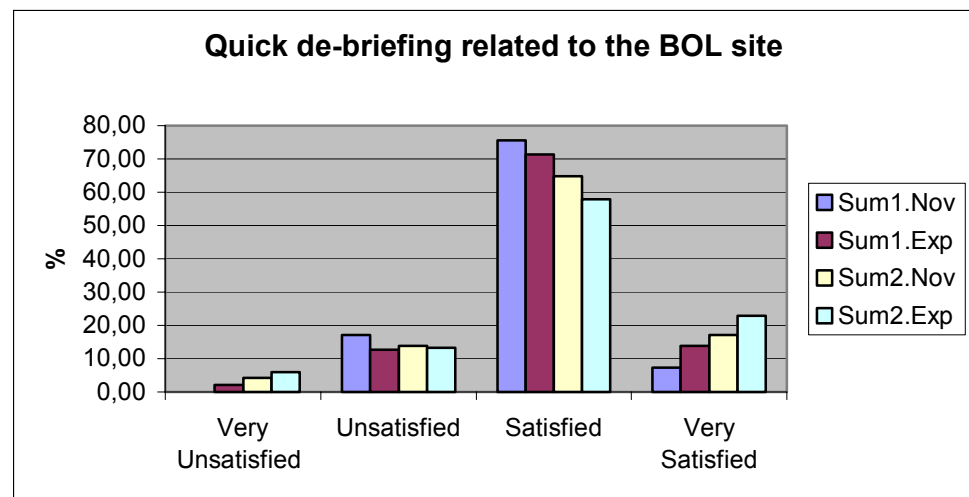


Figure 18: Quick de-briefing related to the BOL site

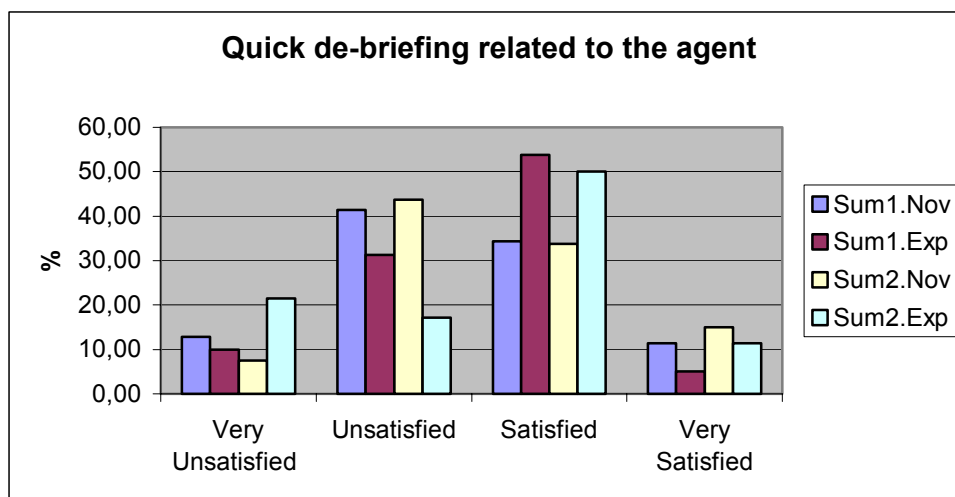


Figure 19: Quick de-briefing related to the agent

However, the interesting part of the evaluation is to consider the more detailed indications given by the test groups concerning their impression and comparison of the two agents.

We have chosen 7 evaluation criteria for the site and the agent seen jointly:

Impression, Command, Effectiveness, Navigability, Learnability, Aidability, and Comprehension, and the description terms are defined in the table below showing the questionnaire scales and definitions.

Scales	Definitions
Impression	The users feelings or emotions when using the software.
Command	The measure to which the user feels that he/she is in control.
Effectiveness	The degree to which the user feels that he/she can complete the task while using the system.
Navigability	The degree to which the user can move around the application
Learnability	The degree to which the user feels that the application is easy to become familiar with.
Aidability	The degree to which the application assists the user to resolve a situation.
Comprehension:	The degree to which the interaction with the application is satisfying.

5.1 The outcome of the evaluation for each scale

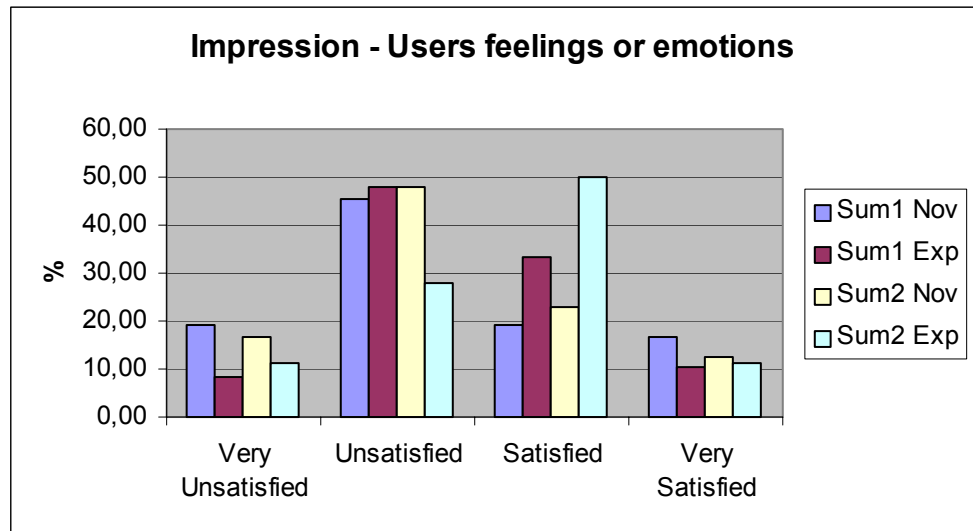


Figure 20 User feelings or emotions

The questions related to the impression of the agent are based on the agent being enjoyable or a bit awkward to use, and if the user would recommend use of the agent to colleagues.

The groups of novices had rather negative feelings for both agents in this respect, probably because novices expect an agent - when being available - should act unimpeachable in all situations. The experienced users, however, are aware of the need of a period for maturing a new product, and in fact, the satisfaction among these users has increased from 44% for the state of the art agent to 61% for the COGITO agent.

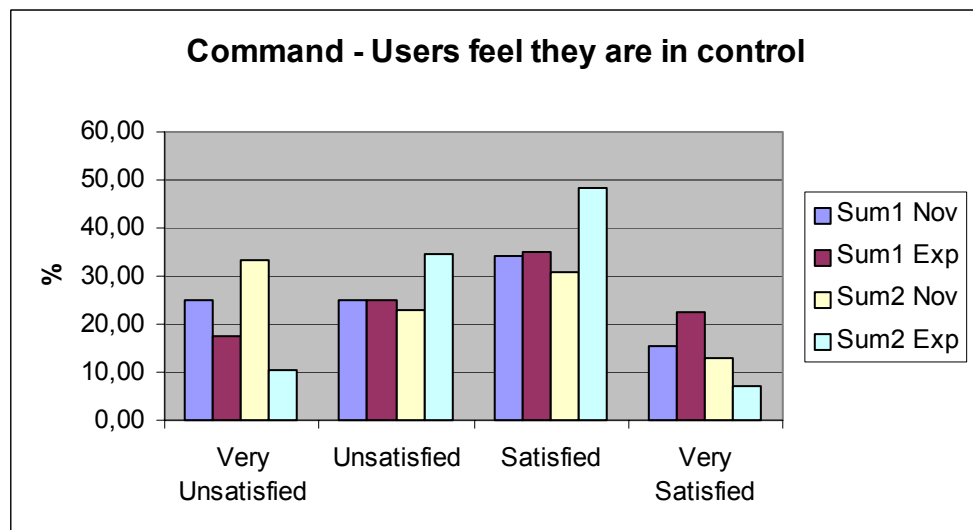


Figure 21: The degree to which users feel in control of.

The questions in this scale are related to the response of the agent concerning start up or re-start of the agent, the speed of the response, and if the agent responded according to expectations.

For this aspect, the overall feelings for both groups of users and for both agents are about 50/50%, which is in fact not bad when testing a new feature. One reason for this rather neutral result may be the frustrating fact valid for both agents that in case the agent 'crashes'

during the session, the user had no real influence on this situation except trying to call up once again. However, when running a remote server this need not be caused by the server itself, but could just as much be caused by the Internet.

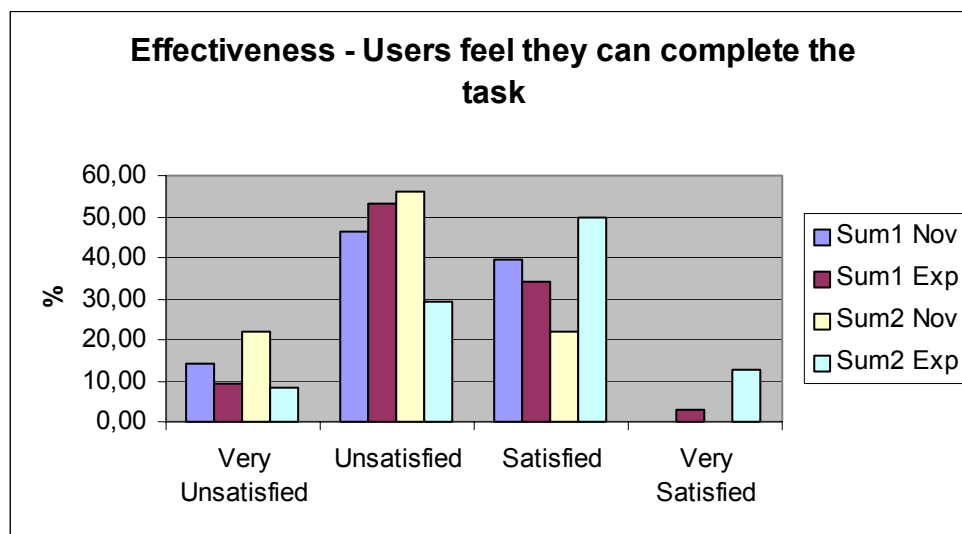


Figure 22: The degree to which users feel they can complete the task.

In this scale, the questions were related to the fulfilment of task, i.e. to find the requested information. Furthermore, the easiness or difficulty in this process is included, exemplified by the number of steps needed to get the information.

For this scale, it seems obvious that the novices do not benefit fully from the features of the agent. The novices testing the COGITO agent seem even more unhappy than the novices testing the state of the art agent. Even though the functionality of the agent and its 'cooperation' or linking to the BOL site was explained to the test persons before each session, this may very well be forgotten during the session for persons not acquainted with the searching procedure using the Internet. So, the novices may have been too much focused on the communication with the agent and, therefore, not been noticing if the requested information was presented through a relevant link to the BOL site (see the section below related to comprehension). For the experience users this feature was observed and utilised, and these users rewarded the more effective linking of the COGITO agent as compared with the state of the art agent by increasing the overall satisfaction from 38% to 63%.

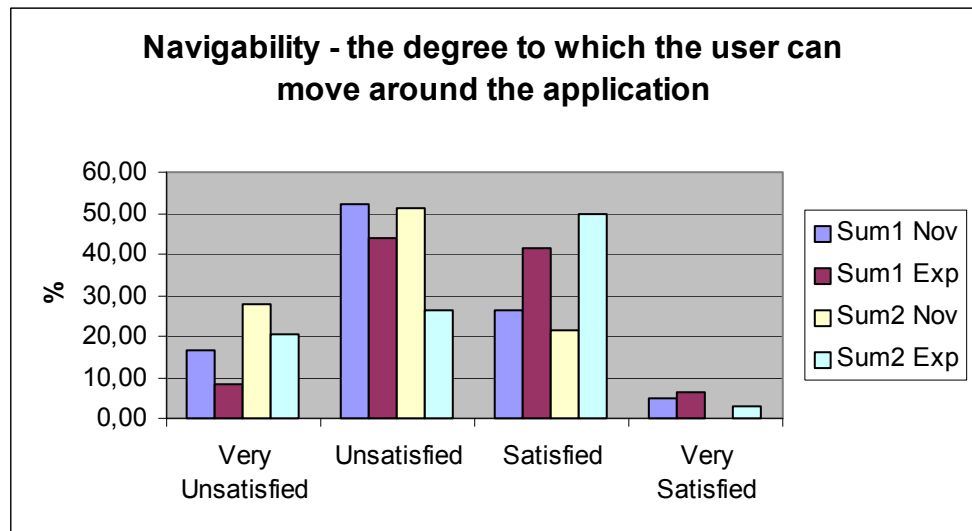


Figure 23 The degree to which the users feel they can navigate using the prototypes

In this scale, the questions were related to the textual support given by the agent concerning how to navigate in order to find the information needed.

The outcome of this scale and the reason for it is very much the same as for the effectiveness even though not to the same degree. The navigability is very much related to the interaction between the agent and the site, and, once again, the experienced users are more aware of and utilise this feature. However, the benefit from the more efficient linking of the COGITO agent as compared with the state of the art agent is in this scale corresponding to an increase in satisfaction of the experienced users of just 3%.

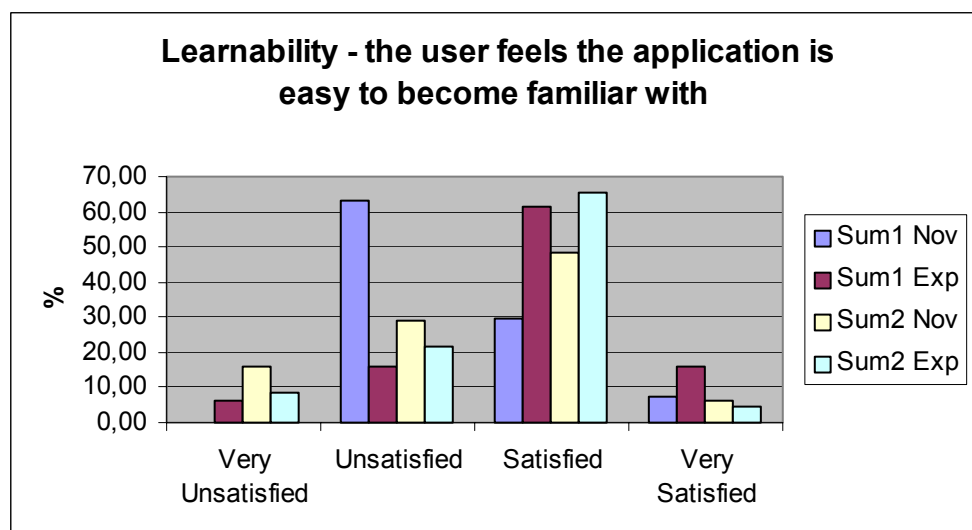


Figure 24 Learnability of the prototypes

The learnability is mostly related to the information needed in beforehand in order to be able to act with the agent, as well as the

acquaintance of the agent based on just short time of experience with the agent.

Concerning the novices, the satisfaction in using the agent increased from 37% for the state of the art agent to 55% for the COGITO agent. The experienced users were even more satisfied concerning learnability of the agent, even though the satisfaction decreased from 77% to 70% going from the state of the art agent to the COGITO agent. The decrease for the experienced users was probably related to the appearance of the agents, which for the state of the art agent was a full torso showing the complete upper part of the body allowing the agent to gesticulate with her arms in order to demonstrate happiness or despair. This kind of agent is more corresponding to the ‘humanlike agent’ suggested in the user requirements. The COGITO agent did not show the arms of the agent, but just allowed her to illustrate her mood by changing the expression of her face. The drastically increase in satisfaction for the novices is related to a ‘mode’ error existing for the state of the art agent, but remedied for the COGITO agent. The mode error for the state of the art agent was due to a tree structure logic of the agent, which did not allow the agent to change from one topic to another without being taken back in the tree structure to the correct starting point. This caused a lot of misunderstandings in the communication between the users and the agent if, e.g., the user jumped from discussing books to making an query about music, and the novices saw this disagreement as a lack of understanding in using the agent.

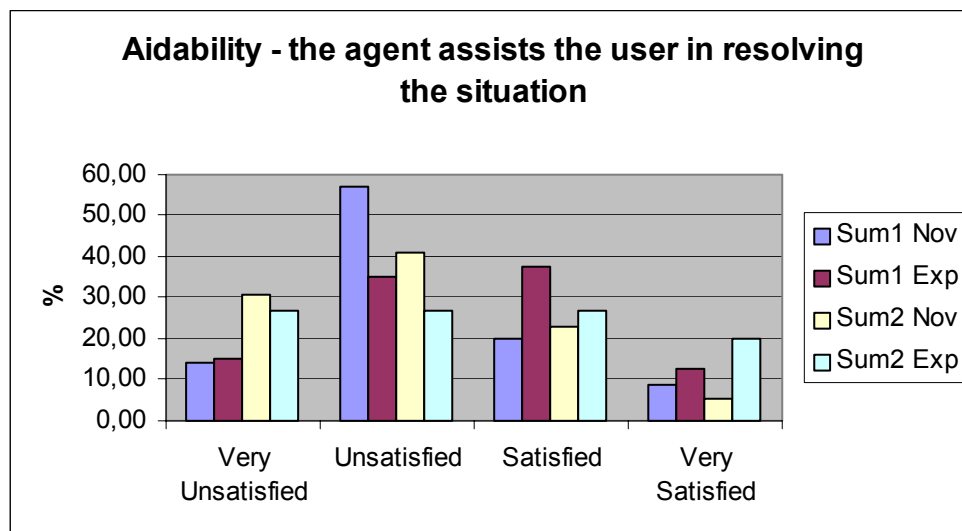


Figure 25: The degree to which the agent assists the user in different situations.

This class deals with the degree to which the agent responded in an instructive way if the communication between the user and the agent did not perform smoothly.

For this class the COGITO agents revealed lack of mature. In case the agent was not able to grasp the meaning of the request from the user, it did not give any informative feed back for a new way of structuring the query. The agent just indicated that ‘this is too much for me, try to formulate the question in another way’ or ‘I don’t understand, what do you mean with this sentence’, but no indication about another or better-

structured ways of communicating. Therefore, in this respect the COGITO agent is very much in line with the state of the art agent judged by the novices as well for as by the experienced users.

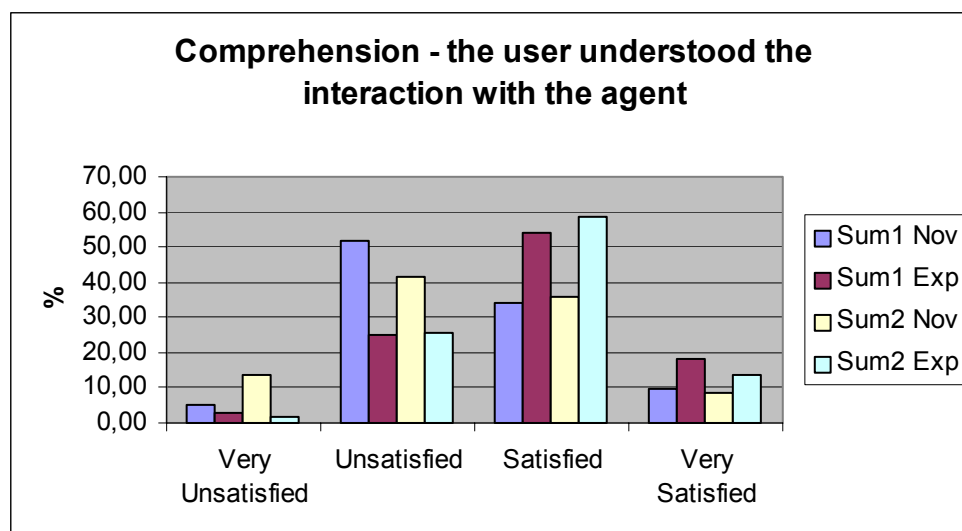


Figure 26: How well did the users understand the interaction between the BOL site and the agent.

In this class, the questions were related to the user's personal feeling of understanding of the information given by the agent, the action of the agent, the interaction between the agent and the BOL site, and about how to operate the agent in relation to this site.

In agreement with the interpretation of the lack of understanding by the novices related to the classes of navigability and effectiveness, the class of comprehension indicates that this is in full agreement with the novices' own feelings. For both agents 44% feel satisfied with their understanding of the interaction, while the feeling of understanding is 72% for the experienced users. Therefore, in total the satisfaction concerning understanding of interaction is more than 50%.

6 Conclusion

The COGITO agent has been evaluated for two groups of test persons, novices and experienced Internet users, to check if the agent is able interactively to facilitate the interaction between the user and an e-commerce site exemplified in the COGITO project by the BOL site offering book, music, and gifts via the Internet.

The COGITO agent developed in the project has been compared with a BOL avatar representing the state of the art of Internet agents of today. The shift in overall satisfaction changing from the BOL avatar to the COGITO agent indicated a slight overall increase in satisfaction for both novices and experienced users (see section 5).

The experienced users seem to appreciate more the developed improvements than the novices. Being more specific, for the topics

selected for the evaluation the satisfaction related to impression of the agent increased for the experienced users from 44% for the state of the art agent to 61% for the COGITO agent. Likewise, the satisfaction related to the effectiveness increased from 38% to 63%. For the novices, however, the satisfaction related to learnability increased from 37% to 55% (see section 5).

The analysis of the conversation logs showed that with an average query length of 5,05 terms the pro-active COGITO agent tend to perform better than traditional web based search engines (e.g., Excite) without demanding the users to use any Boolean operators, and with good performance in terms of search results. In addition, the pro-activeness of the COGITO agent seems to urge people to type in more words than the more "passive" BOL prototype where users tend to use the agent as a traditional search machine. The COGITO queries on an average contained more than twice the number of terms (app. 120% more) compared with the averages on traditional search machines and more than 70% more than the BOL users (see section 3).

Another important result, with respect to increased performance of the COGITO agent compared to the BOL agent, were the reduction in mode errors, i.e. the ability to shift easily from one topic to another, and a proactive behaviour in which the agent utilises the query from the user in producing a more proactive behaviour of communication (see section 3). In addition, The COGITO agent shows a better performance (61% vs. 47%) than the BOL agent with respect to correct feedback to user input. It seems that the COGITO agent is better in a pro-active way to recognize search terms.

The analysis of the eye-tracking data showed that more visual attention has been paid to the BOL agent visualisation compared to the animation of the COGITO agent. The BOL agent animation has attracted approximately double as much visual attention than the COGITO agent. This is probably due to the more the photo-like appearance, the obliging attitude, and a larger repertoire of gesticulations compared to the cartoon-like appearance of the COGITO agent. The result should be seen in relation to the small amount of viewing time spent on the agent visualisations. For both prototypes, an equal amount of visual attention was paid to the agent as a whole and the BOL site. Moreover, the eye-movement analysis showed that approximately half of the users visual attention has been paid to the keyboard plus the agents' input field while typing.

The COGITO agent evaluated was the version prepared for late December 2001, but for the remaining two months the agent has been further improved by implementing additional features to be demonstrated during the COGITO review start of March 2002.

7 References

- Andersen, V.; Hansen, C.B.; Andersen, H.H.K., *Evaluation of agents and study of end-user needs and behaviour for e-commerce. COGITO focus group experiment. Risø-R-1264(EN)*, 2001
- Fenichel, C. H., (1981). Online searching: Measures that discriminate among users with different types of experience. *Journal of the American Society for Information Science*, 32, 23-32.
- Jansen, Bernard J., Spink, Amanda, & Saracevic, Tefko (2000). Real Life, Real Users, and Real Needs: A Study and Analysis of User Queries on the Web. *Information Processing & Management*, 36, 207-227.
- Spink, A., Saracevic, T. (1997). Interactive information retrieval: Sources and effectiveness of search terms during mediated online searching. *Journal of the American Society for Information Science*, 48(8), 741-761.

Appendices

Appendix 1: List of tasks to be solved by the groups of test persons using the BOL avatar

- 1 Bevor Sie ein Buch bei BOL kaufen, wollen Sie ein wenig über die Buchhandlung wissen, z.B. wieviele Bücher auf Lager sind und was für Bücher angeboten werden (z.B. Belletristik, Fachliteratur, usw.) Versuchen Sie eine Übersicht über die Bücher zu bekommen und lassen Sie sich von dem Agenten helfen.
- 2 Sie suchen ein Geschenk für einen Freund, aber Sie sind sich noch nicht sicher darüber was es sein soll. Benutzen Sie den Agenten, um mögliche Vorschläge zu finden.
- 3 Sie haben den Titel von einem Buch. Finden sie es auf der Site und untersuchen Sie den Preis des Buches, wie Sie es geliefert bekommen, wie lang die Lieferzeit ist, und was die Lieferung kostet. Evt. suchen Sie das Buch 'Der Hobbit'.
- 4 Da es sich um ein Geschenk handelt, ist es wichtig zu wissen, ob das Buch umgetauscht werden kann und zu welchen Bedingungen. Probieren Sie, ob der Agent hierbei helfen kann.
- 5 Was außer Büchern kann man sonst noch kaufen?
- 6 Eines der Kinder in Ihrer Familie bekommt eine Playstation zu Weihnachten. Sie wollen gerne dem Kind ein Spiel dazu schenken.
- 7 Sie haben einen Bekannten , der gerne Bücher auf englisch liest. Sie wollen untersuchen, ob man englisch geschriebene Bücher bei BOL kaufen kann.
- 8 Sie sind neugierig herauszufinden, wie es ist elektronische Bücher zu lesen. Lassen Sie sich von dem Agenten die unterschiedlichen Möglichkeiten zeigen.
- 9 Sie möchten gerne wissen, welche Fragen von anderen Kunden in Verbindung mit dem Handel bei BOL gestellt werden. Kann der Agent Ihnen bei solchen Auskünften helfen?
- 10 Stellen Sie sich vor Sie besuchen BOL zum ersten mal. Sie möchten gerne eine Demonstration der Site sehen. Kann der Agent Ihnen dabei helfen?
- 11 Sie möchten gerne über die unterschiedlichen Suchmethoden auf der Site unterrichtet werden. Ist es möglich eine Übersicht über die unterschiedlichen Möglichkeiten zu bekommen?

- 12 Sie sind nervös darüber in eine unübersichtliche Situation zu geraten und möchten gerne die Sicherheit haben mit Personen von BOL in Kontakt zu kommen, die Ihnen weiterhelfen könnten. Ist dies möglich?
- 13 Sie sind neugierig zu wissen, wie man bei einer Bestellung verfährt. Ist es möglich dies beschrieben zu bekommen?
- 14 Schon seit langem steht auf Ihrer Wunschliste ein Fahrradurlaub. Sie möchten gerne Literatur über verschiedene Routen und mit praktischen Rat, wie es ist mit dem Fahrrad unterwegs zu sein, finden.
- 15 Sie haben ein Buch als Geschenk für einen Bekannten gefunden, möchten aber gerne wissen was andere Käufer über das Buch gemeint haben?
- 16 Sie haben einen guten Bekannten, der gerne ein Lexikon für seine Freundin kaufen will. Es soll schön aussehen, aber nicht zu teuer sein. Welche Angebote sind vorhanden?

Appendix 2: List of tasks to be solved by the groups of test persons using the COGITO agent

1. Bevor Sie ein Buch bei BOL kaufen, möchten Sie zunächst ein wenig über die Buchhandlung wissen, z.B. wieviele Bücher auf Lager sind und was für Bücher angeboten werden (z.B. Belletristik, Fachliteratur, usw.) Versuchen Sie eine Übersicht über die Bücher zu bekommen und lassen Sie sich von dem Agenten helfen.
2. Sie kennen den Autor eines Handbuchs über Computer. Finden Sie das Buch mit Hilfe des Agenten und finden Sie den Preis des Buches heraus, wie Sie es geliefert bekommen, wie lang die Lieferzeit ist, und was die Lieferung kostet. Sie könnten beispielsweise ein Buch von Beate Majetschak suchen. Gibt es Kundenmeinungen zu diesem Buch?
3. Finden Sie heraus, wie man ein Buch bestellen kann, wie man es bezahlen kann und wie es mit den Sicherheitsvorkehrungen bei der Bezahlung aussieht.
4. Sie haben einen Bekannten, der gerne Bücher auf englisch liest. Sie wollen herausfinden, ob man auf englisch geschriebene Bücher bei BOL kaufen kann.
5. Sie möchten mehr über den Agenten wissen. Was würden Sie den Agenten fragen? Probieren Sie es aus!
6. Versuchen Sie, mit dem Agenten zu “chatten” und beobachten Sie, was dabei auf der BOL-Website passiert. Sie könnten beispielsweise über Internet Entwicklungen sprechen.
7. Stellen Sie sich vor, Sie als Amateur wären sehr daran interessiert, etwas im Web zu erstellen, besonders im Bereich der “Bildbearbeitung”. Kann der Agent Ihnen dabei helfen und falls ja, wählen Sie ein günstiges Programm samt Benutzerhandbuch aus der Auswahl aus.
8. Sie sind ein guter Web-Programmierer. Benutzen Sie den Agenten, um Bücher über Programmier-Sprachen zu finden (stellen Sie sich vor, java oder html wären Ihre Lieblings-Programme). Welche Vorschläge macht Ihnen der Agent? Finden Sie etwas über mögliche Kundenkommentare zu einem der Vorschläge heraus. Versuchen Sie, die Fragen des Agenten zu beantworten.
9. Sie haben einen Teil-Zeit-Job als Assistenz-Lehrer in Mathematik an einer öffentlichen Schule. Sie hatten bereits gute Erfahrungen damit, “Lernsoftware” zu benutzen, wenn es darum ging die Schüler in diesem Fach zu beeindrucken. Allerdings besitzt die Schule keine

„Lernsoftware“. Der Direktor hat aber beschlossen, eine solche Software zu kaufen, und Sie sind nun dafür verantwortlich, ein gutes Angebot zu finden, nicht nur für Mathematik, sondern auch für Deutsch und Informatik. Kann der Agent Ihnen helfen?

10. Da Sie Ihre Sache bei der Suche nach einer „Lernsoftware“ für die Schüler gut gemacht haben, findet der Direktor, Sie seien die richtige Person, um die Computer-Ausstattung der Schule zu aktualisieren. Vor allem möchte die Schule ihr Betriebssystem auf Windows 2000 Pro umstellen und das Office Paket für Word und Tabellenbearbeitung implementieren. Versuchen Sie, mit dem Agenten darüber zu diskutieren.
11. Der Direktor ist sehr zufrieden mit Ihnen und bietet Ihnen eine Position als Festangestellte/r mit besonderer Verantwortung für die Informatik. Allerdings sind die älteren Lehrer ein wenig skeptisch bzgl. dieser neuen Initiativen. Vor allem finden sie, dass die Handbücher für Windows 2000 und das Office Paket ihnen nicht ausreichend helfen, die neue Software kennen zu lernen. Sie kennen die Mängel der Handbücher und suchen nun einfach zugängliche Hefte, die nicht zu teuer sind. Kann Ihnen der Agent hierbei helfen?

Appendix 3: Background information of the test persons

Hintergrundinformationen**Datum:** _____

Geschlecht: m/w _____

Alter: _____

Ausbildung: _____

Beruf /Position: _____

Seit wie vielen Jahren benutzen Sie einen Computer?

- _____ Mehr als 5 Jahre
_____ Zwischen 2 und 5 Jahre
_____ Weniger als 2 Jahre

Benutzen Sie das Internet?

Am Arbeitsplatz:

- _____ Nein
_____ Ja, seit dem Jahr: _____

Zu Hause:

- _____ Nein
_____ Ja, seit dem Jahr: _____

Wie häufig benutzen Sie das Internet?

Am Arbeitsplatz:

- _____ Jeden Tag
_____ Häufig, z.B. mehrmals die Woche
_____ Selten, z.B. mehrmals im Monat

Zu Hause:

- _____ Jeden Tag
_____ Häufig, z.B. mehrmals die Woche
_____ Selten, z.B. mehrmals im Monat

Wozu benutzen Sie das Internet?

- | | |
|-----------------------------|---------------------------------|
| _____ E-mail | _____ Homebanking |
| _____ Informationssuche | _____ Musik |
| _____ Web-design | _____ Spiele |
| _____ Dokumentendatenbanken | _____ sonstiges, bitte angeben: |

(Diskussionsforen, Filmdatenbanken, etc.)

Haben Sie über das Internet eingekauft?

Appendix 4: Questionnaire for the impression of the test persons related to the COGITO agent/BOL site

Fragebogen

Testperson: _____

Affektive (gefühlsbetonte), hedonische (sinnliche) Qualität
Markieren Sie die Eindrücke, die zu diesem Zeitpunkt Ihre Gefühle über die Benutzung der BOL-site mit den Agenten am besten beschreiben

Fragen zur BOL-site ohne Berücksichtigung des Agenten:
(In der folgenden Tabelle kreuzen Sie bitte "Ja" oder "Nein" an)

	Ja	Nein		Ja	Nein
Wie erwartet			unterhaltend/ anregend		
motivierend, herausfordernd			ermöglicht neue Erfahrungen		
nichts neues			ein bibchen überwältigend		
unpassend			eindrucksvoll		
frustrierend, schockierend			kann nicht genug davon bekommen		

Fragen zum **BOL Agenten**:
(In der folgenden Tabelle kreuzen Sie bitte "Ja" oder "Nein" an)

	Ja	Nein		Ja	Nein
Wie erwartet			frustrierend, schockierend		
schaftt Unsi- cherheit			unpersönlich		
aufdringlich			unterhaltsam		
keine grobe Sache/ macht kaum einen Unterschied			ermöglicht neue Erfahrungen		
anregend			spart Zeit		
mühsam, zeitraubend			(potenziell) gewohnheitsbilde nd		
einnehmend			hilft am Anfang bei Schwierigkeiten		
motivierend, herausfordernd			ein bibchen überwältigend		
nichts neues			brauchbare Eigenschaft/Funkt ion		
zu modern / verfrühte Anwendung			verwirrend		
Zeitverschwen dung			eindrucksvoll		

unpassend			kann genug bekommen	nicht davon		
-----------	--	--	---------------------------	----------------	--	--

Kurzer Fragebogen zur Nachbereitung

Fragen zur BOL-site **ohne Berücksichtigung des Agenten.**

(Beachten Sie bitte die unterschiedliche Reihenfolge der Beurteilungskriterien von Frage zu Frage!!)

Bestimmte Informatione n leicht zu finden?	Sehr befrie digen d	Befriedigend	Unbefriedigend	Sehr unbefrie digend
Lesbarkeit des Inhalts?	Sehr unbef riedig end	Unbefriedigend	Befriedigend	Sehr befriedig end
Logik im navigieren?	Sehr unbef riedig end	Unbefriedigend	Befriedigend	Sehr befriedig end
Erscheinung der Site?	Sehr befrie digen d	Befriedigend	Unbefriedigend	Sehr unbefrie digend
Qualität der Graphik?	Sehr befrie digen d	Befriedigend	Unbefriedigend	Sehr unbefrie digend
Stärkung der Marke BOL?	Sehr unbef riedig end	Unbefriedigend	Befriedigend	Sehr befriedig end
Geschwindi gkeit der Site?	Sehr befrie digen d	Befriedigend	Unbefriedigend	Sehr unbefrie digend
Macht die Benutzung Spab?	Sehr befrie digen d	Befriedigend	Unbefriedigend	Sehr unbefrie digend
Wie sind die Erklärungen zur Benutzung der Site?	Sehr unbef riedig end	Unbefriedigend	Befriedigend	Sehr befriedig end
Gesamteind ruck zur einfachen	Sehr unbef riedig	Unbefriedigend	Befriedigend	Sehr befriedig end

Benutzung?	end			
Ihr Gesamteindruck zum Umgang mit der Site?	Sehr befriedigend	Befriedigend	Unbefriedigend	Sehr unbefriedigend
Eine 'coole' Site?	Sehr unbefriedigend	Unbefriedigend	Befriedigend	Sehr befriedigend

Fragen zum BOL-Agenten:

Bestimmte Informationen sind leicht zu bekommen?	Sehr befriedigend	Befriedigend	Unbefriedigend	Sehr unbefriedigend
Die Erscheinung des Agenten?	Sehr befriedigend	Befriedigend	Unbefriedigend	Sehr unbefriedigend
Qualität der Graphik?	Sehr befriedigend	Befriedigend	Unbefriedigend	Sehr unbefriedigend
Stärkung der Marke BOL?	Sehr unbefriedigend	Unbefriedigend	Befriedigend	Sehr befriedigend
Geschwindigkeit der Site?	Sehr befriedigend	Befriedigend	Unbefriedigend	Sehr unbefriedigend
Macht die Benutzung Spab?	Sehr befriedigend	Befriedigend	Unbefriedigend	Sehr unbefriedigend
Wie sind die Erklärungen zur Benutzung der Site?	Sehr unbefriedigend	Unbefriedigend	Befriedigend	Sehr befriedigend
Gesamteindruck zur einfachen Benutzung?	Sehr unbefriedigend	Unbefriedigend	Befriedigend	Sehr befriedigend
Ihr Gesamteindruck zum	Sehr befriedigend	Befriedigend	Unbefriedigend	Sehr unbefriedigend

Umgang mit der Site?	d			
Ein 'cooler' Agent?	Sehr unbef riedig end	Unbefriedigend	Befriedigend	Sehr befriedige nd

Detaillierter Fragebogen zur Nachbereitung

Wir haben 7 Evaluierungskriterien für die Site und den Agenten ausgewählt. Dies sind: Eindruck, Kontrolle, Effektivität, Navigation, Lernfähigkeit, Hilfestellung und Verständnis. Diese Begriffe sind in der nächsten Tabelle definiert.

Begriff	Definition
Eindruck	Die Gefühle und Gemütsbewegungen des Benutzers bei der Benutzung
Kontrolle	Inwieweit der Benutzer fühlt, daß er/sie die Kontrolle hat
Effektivität	Inwieweit der Benutzer während der Benutzung des Systems fühlt, daß er/sie die Aufgaben zufriedenstellend ausführen kann
Navigierbarkeit	Inwieweit der Benutzer sich in der Applikation bewegen kann
Lernbarkeit	Inwieweit der Benutzer fühlt, daß er mit der Applikation schnell vertraut wird
Hilfestellung	Inwieweit die Applikation dem Benutzer hilft, Probleme zu lösen
Verständnis	Inwieweit der Benutzer die Interaktion mit der Applikation versteht.

Die Liste mit Fragen die in der detaillierten Nachbereitung benutzt wird

Eindruck- die Gefühle und Gemütsbewegungen des Benutzers bei der Benutzung			
Die BOL site (mit Agent) ist eine Site, die ich regelmäßig benützen würde.	Stimme sehr zu	Stimme zu	Stimme nicht zu
Die BOL site (mit Agent) war sehr anspruchsvoll und nicht sehr angenehm in der Benutzung	Stimme sehr zu	Stimme zu	Stimme nicht zu
Dieser hoch-technologische Weg zur Informationsfindung ist einschüchternd	Stimme überhaupt nicht zu	Stimme nicht zu	Stimme zu
Die Benutzung des Agenten war frustrierend	immer	meistens	Teilweise
Ich würde diese BOL site (mit Agent) meinen	Stimme sehr zu	Stimme zu	Stimme nicht zu

Kollegen empfehlen				
Mir hat die Arbeit mit diesem Agenten gefallen	Immer	Meistens		Teilweise
Diese BOL site (mit Agent) war wirklich sehr schwierig zu benutzen	Immer	Meistens		Teilweise
Kontrolle – inwieweit der Benutzer fühlt, daß er die Kontrolle hat				
Der Agent hat einfach und schnell auf meine Eingaben reagiert	Immer	Meistens	Teilweise	selten
Ich fand es nicht einfach, die Applikation zu starten	Stimme sehr zu	Stimme zu	Stimme nicht zu	Stimme überhaupt nicht zu
Ich fühle, dass ich keine Kontrolle über den Agenten hatte	Stimme überhaupt nicht zu	Stimme nicht zu	Stimme zu	Stimme sehr zu
Der Agent hat nicht schnell genug auf meine Eingaben reagiert	Stimme überhaupt nicht zu	Stimme nicht zu	Stimme zu	Stimme sehr zu
In den Fällen wo der Agent aufgehört hat zu arbeiten, war es nicht leicht ihn erneut zu starten	Stimme sehr zu	Stimme zu	Stimme nicht zu	Er hörte nicht auf
Es war einfach, den Agenten exakt nach meinen Wünschen zu steuern	selten	Teilweise	Meistens	Immer
Effektivität – Inwieweit der Benutzer während der Benutzung des Systems fühlt, daß er/sie die Aufgaben zufriedenstellend ausführen kann				
Die Benutzung der Site, um die gewünschte Information zu bekommen, war einfach	Immer	Meistens	Teilweise	selten

Die Benutzung des Agenten behinderte mich in der Bearbeitung der Aufgabe	Stimme sehr zu	Stimme zu	Stimme nicht zu	Stimme überhaupt nicht zu
Ich habe die gewünschte Information nicht gefunden	Immer	Meistens	Teilweise	selten
Es werden zu viele Schritte benötigt, um die Information zu bekommen	Stimme überhaupt nicht zu	Stimme nicht zu	Stimme zu	Stimme sehr zu
Navigation – Inwieweit der Benutzer sich in der Applikation bewegen kann				
Es war leicht sich zurecht zu finden, wenn man die Informationen des Agenten befolgte	Stimme sehr zu	Stimme zu	Stimme nicht zu	Stimme überhaupt nicht zu
Es gab viele Möglichkeiten, die gewünschte Information zu finden	Stimme überhaupt nicht zu	Stimme nicht zu	Stimme zu	Stimme sehr zu
Der Agent lieferte brauchbare “short cuts” (Abkürzungswege)	Stimme sehr zu	Stimme zu	Stimme nicht zu	Stimme überhaupt nicht zu
Mit Hilfe des Agenten konnte ich meine Information finden	Immer	Meistens	Teilweise	selten
Bei der Benutzung des Agenten ist man schnell desorientiert	Stimme sehr zu	Stimme zu	Stimme nicht zu	Stimme überhaupt nicht zu
Ich wusste wohin der nächste Schritt mit dem Agenten führen sollte	selten	Teilweise	Meistens	Immer
Lernbarkeit – Inwieweit der Benutzer fühlt, dass er mit der Applikation schnell vertraut wird				
Ich würde mich sicherer fühlen,	Stimme sehr zu	Stimme zu	Stimme nicht zu	Stimme überhaupt

wenn ich nur die Methoden zur Informationssuche benutzen könnte, mit denen ich schon vertraut bin.				nicht zu
Die einleitenden Erklärungen, wie der Agent zu benutzen ist, waren gut genug, um ihn anzuwenden.	Stimme überhaupt nicht zu	Stimme nicht zu	Stimme zu	Stimme sehr zu
Ich fühlte mich sicher als ich die verschiedenen Möglichkeiten zur Informationsfindung ausprobierte	Stimme sehr zu	Stimme zu	Stimme nicht zu	Stimme überhaupt nicht zu
Die Benutzung des Agenten war leicht zu lernen	Sehr einfach	einfach	Nicht so einfach	schwierig
Hilfestellung Inwieweit die Applikation dem Benutzer hilft, Probleme zu lösen				
Der Agent hat mir geholfen, Fehler zu vermeiden	Immer	Meistens	Teilweise	selten
Die angebotene Hilfe vom Agenten gibt genug Information	Stimme überhaupt nicht zu	Stimme nicht zu	Stimme zu	Stimme sehr zu
Ich war unsicher, ob ich in der richtigen Art und Weise zu dem Agenten "gesprochen" habe	Immer	Meistens	Teilweise	selten
Der Agent gab nicht genug Information, wie man verfahren/fortfahren sollte	Immer	Meistens	Teilweise	selten
Ich fand, die neue BOL site war schwierig zu benutzen, wenn ich etwas ohne den Agenten versuchte	Stimme überhaupt nicht zu	Stimme nicht zu	Stimme zu	Stimme sehr zu
Ich konnte die	Stimme	Stimme	Stimme	Stimme

gelieferte Information von dieser neuen BOL site nicht verstehen oder sie anwenden.	überhaupt nicht zu	nicht zu	zu	sehr zu
Verständnis	Inwieweit der Benutzer die Interaktion mit der Applikation versteht.			
Die vom Agenten angebotene Information war klar /eindeutig präsentiert und zu verstehen	Stimme sehr zu	Stimme zu	Stimme nicht zu	Stimme überhaupt nicht zu
Die von dem Agenten gezeigten Informationen waren nicht konsistent	Stimme sehr zu	Stimme zu	Stimme nicht zu	Stimme überhaupt nicht zu
Die mit dem Agenten verbundenen Aktionen waren leicht zu verstehen		Stimme nicht zu	Stimme zu	Stimme sehr zu
Ich habe die Interaktion zwischen Agenten und BOL site verstanden	Immer	Meistens	Teilweise	selten
Ich habe verstanden, wie man den Agenten auf der BOL Site anwendet	Immer	Meistens	Teilweise	selten
Um die gesuchte Information zu finden, mußte ich öfters fragen	Immer	Meistens	Teilweise	selten
Das Layout der Handbücher war zufriedenstellend	Stimme sehr zu	Stimme zu	Stimme nicht zu	Stimme überhaupt nicht zu
Die Animationen des Agenten waren klar und deutlich	Stimme sehr zu	Stimme zu	Stimme nicht zu	Stimme überhaupt nicht zu
Die Antworten des Agenten waren zu	Stimme überhaupt	Stimme nicht zu	Stimme zu	Stimme sehr zu

komplex	nicht zu			
Ich habe die Fragen des Agenten komplett verstanden	Stimme überhaupt nicht zu	Stimme nicht zu	Stimme zu	Stimme sehr zu

Appendix 5: Introduction to users

Untersuchung

- Kooperationsprojekt mit Internetshop bol
- Es geht um die Bewertung von bol.de und eines neuen virtuellen Assistenten/Agenten.
- Wir werden einige Situationen durchspielen (keine „Aufgaben“).
- Dabei sollen Sie laut aussprechen, was Sie denken. (Das ist vielleicht ungewohnt.)
- Außerdem nehmen wir die Untersuchung auf Video auf.
- Einverständniserklärung unterschreiben
- Schließlich registrieren wir Ihre Blickbewegungen.
- Erklärung der Komponenten
- Herr Andersen bleibt im Raum und bedient das Gerät.
- Nach ca. 30 Minuten sind wir hier fertig, dann füllen Sie noch einen Fragebogen aus.
- Zum Schluss erhalten Sie DM 60,- (€ 30,-).
- Nicht Sie werden getestet, sondern der Agent!
- Kalibrieren
- Bis an den Tischrand
- Hinten anlehnen
- Möglichst nicht bewegen
- Gleich erscheint in der Mitte ein Punkt
- Bitte fixieren
- Der Punkt wandert um den Bildschirm
- Bitte mit dem Blick folgen, dabei NUR DIE AUGEN BEWEGEN
- Ab jetzt bitte die nächsten 20 Minuten möglichst nicht bewegen
- Lautes Denken
- Bitte laut vorlesen und dann kommentieren, was Sie tun und warum Sie genau dies tun. Lassen Sie uns an Ihren Gedanken teilhaben. (Wie ein Radioreporter)

Bibliographic Data Sheet**Risø-R-1363 (EN)**

Title and authors

Evaluation of the COGITO system

Verner Andersen and Hans .H.K. Andersen

ISBN

87-550-3111-0 (Internet)

ISSN

0106-2840

Department or group

System Analysis Department, SPM

Date

August 2002

Groups own reg. number(s)

Project/contract No(s)

IST-1999-13347

Sponsorship

Pages

54

Tables

7

Illustrations

26

References

4

Abstract (max. 2000 characters)

This report covers the deliverable D7.2 of the COGITO project. It presents the evaluation of an 'intelligent' agent integrated into an e-commerce system. The aim of the agent is to support the user partly through direct communication and partly by guiding him/her for navigating the web-site and for benefiting efficiently from the functionality of the site. In order to make the report a stand-alone report, it contains an introduction including 'Study of end-user needs and behaviour' exposing problems concerning e-commerce in general (taken from the D6.1), and a brief review of the objectives for the COGITO project in order to clarify the aims of the project. However, the important part of the report presents the methodology for evaluating the performance of the COGITO agent and the outcome of this evaluation. The questionnaire for eliciting the background information for the test persons and their personal responses to the performance of the agent is presented in appendices

Descriptors INIS/EDB

Usability Evaluation, Intelligent Systems, Software Agents.